

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:02:34 ; Search time 2.37192 Seconds
(without alignments)
620.235 Million cell updates/sec

Title: us-09-863-901-22

Perfect score: 258

Sequence: 1 SGGSGGGSGSGSGGQSG.....SGSGGGSGSGGGGGLRS 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	65.9	738	3	US-08-864-038A-3
2	154	59.7	235	2	US-08-529-190B-1
3	154	59.7	641	4	US-09-249-585A-3
4	147	57.0	201	4	US-09-052-995-1
5	147	57.0	201	4	US-09-053-003-40
6	147	57.0	201	4	US-09-054-281-22
7	147	57.0	201	4	US-09-478-948-6
8	140	54.3	40	3	US-08-815-190A-17
9	134.5	52.1	907	2	US-09-010-928B-4
10	129	50.0	647	2	US-08-770-761A-8
11	129	50.0	705	2	US-08-770-761A-7
12	124.5	48.3	344	1	US-08-891-254-7
13	124.5	48.3	344	2	US-08-819-539-7
14	124.5	48.3	344	2	US-09-030-270A-7
15	124.5	48.3	344	4	US-08-984-207-7
16	124.5	48.3	344	4	US-09-013-587-7
17	124.5	48.3	344	5	PCT-US96-08819-7
18	124	48.1	322	1	US-08-014-943A-2
19	124	48.1	322	1	US-08-486-421-3
20	124	48.1	322	1	US-08-470-911-3
21	124	48.1	322	2	US-08-486-809-3
22	123	47.7	49	3	US-08-469-318-194
23	123	47.7	49	3	US-08-468-609A-194
24	123	47.7	49	4	US-08-446-872A-194
25	123	47.7	49	4	US-08-762-227A-194
26	123	47.7	49	5	PCT-US95-01185-194
27	123	47.7	335	3	US-08-469-318-143

Sequence 143, App
Sequence 143, App
Sequence 143, App
Sequence 143, App
Sequence 5, Appli
Sequence 4, Appli
Sequence 195, App
Sequence 195, App
Sequence 195, App
Sequence 195, App
Sequence 195, App
Sequence 34, Appli
Sequence 16, Appli
Sequence 16, Appli
Sequence 16, Appli
Sequence 34, Appli
Sequence 16, Appli

28 123 47.7 335 3 US-08-468-609A-143
29 123 47.7 335 4 US-08-446-872A-143
30 123 47.7 335 4 US-08-762-227A-143
31 123 47.7 335 5 PCT-US95-01185-143
32 123 47.7 447 4 US-09-916-109-5
33 123 47.7 483 4 US-09-916-109-4
34 122 47.3 60 3 US-08-469-318-195
35 122 47.3 60 3 US-08-468-609A-195
36 122 47.3 60 4 US-08-446-872A-195
37 122 47.3 60 4 US-08-762-227A-195
38 122 47.3 60 5 PCT-US95-01185-195
39 122 47.3 211 1 US-08-276-852-34
40 122 47.3 211 1 US-08-133-011-16
41 122 47.3 211 1 US-08-322-730A-16
42 122 47.3 211 1 US-08-387-874-16
43 122 47.3 211 1 US-08-899-575-34
44 122 47.3 211 1 US-08-899-575-34
45 122 47.3 211 2 US-08-383-619-16

ALIGNMENTS

RESULT 1

US-08-864-038A-3

; Sequence 3, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: 812-5 Hirano

; STREET: Isshinden

; CITY: Tsu-city

; STATE: Mie-Prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 MB storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Microsoft Windows 95

; SOFTWARE: Word Perfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A

; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-184459

; FILING DATE: 15-July-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: C. Bruce Hamburg

; REGISTRATION NUMBER: 22,389

; REFERENCE/DOCKET NUMBER: F-5610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)986-2340

; TELEFAX: (212)953-7733

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 738

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Pinctada fucata

; CELL TYPE: mantle epithelial cell

; FEATURE:

; NAME/KEY: peptide

; LOCATION: from 1 to 738

; IDENTIFICATION METHOD: E (by experiment)

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,281
FILING DATE: 02-APR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/080,234
FILING DATE: 03-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/826,622
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Timothy L.
REGISTRATION NUMBER: 35,367
REFERENCE/DOCKET NUMBER: 015280-295200US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..97
OTHER INFORMATION: /product= "OTHER"
OTHER INFORMATION: /note= "Gly at positions 1-97 may be
OTHER INFORMATION: present or absent"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 105..201
OTHER INFORMATION: /product= "OTHER"
OTHER INFORMATION: /note= "Gly at positions 105-201 may be
OTHER INFORMATION: present or absent"
US-09-054-281-22

Query Match 57.0%; Score 147; DB 4; Length 201;
Best Local Similarity 56.5%; Pred. No. 1.2e-06;
Matches 26; Conservative 0; Mismatches 20; Indels 0; G

RESULT 7
 US-09-478-948-6
 : Sequence 6, Application US/09478948
 47
 2
 58
 Db

```

? GENERAL INFORMATION:
? APPLICANT: Shan, Bei
? APPLICANT: Okamoto, Arthur Y.
? APPLICANT: Tularik, Inc.
? APPLICANT: Tularik
? TITLE OF INVENTION: FXR Receptor-Mediated Modulation of Cholesterol
? TITLE OF INVENTION: Metabolism
? FILE REFERENCE: 018781-001310US
? CURRENT APPLICATION NUMBER: US/09/478,948
? CURRENT FILING DATE: 2000-01-06
? PRIOR APPLICATION NUMBER: US 60/115,249
? PRIOR FILING DATE: 1999-01-07

```

[illegible]

[illegible]

```

CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,270A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,226
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1521
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-030-270A-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;
Best Local Similarity 40.6%; Pred. NO. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps

QY 3 GSGSGGGGSGGSGGQ-----SGSGSGGGGSGG-SGSGGGGSGGSGG 43
Db 141 GNGVGANGAGKAGGCGGGGLAEALQETQIQLGCGGAGAGGGVGGAGGADGGSGAG 200

QY 44 GQSG 47
Db 201 GAGG 204

RESULT 15
US-08-984-207-7
: Sequence 7, Application US/08984207
: Patent No. 6235974
: GENERAL INFORMATION:
: APPLICANT: Oiu, Dewen
: APPLICANT: Wei, Zhong-Min
: APPLICANT: Beer, Steven V.
: TITLE OF INVENTION: HIPERSENSITIVE RESPONSE INDUCED
: TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
: STREET: P. O. Box 1051, Clinton Square
: CITY: Rochester
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/984,207
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:

```

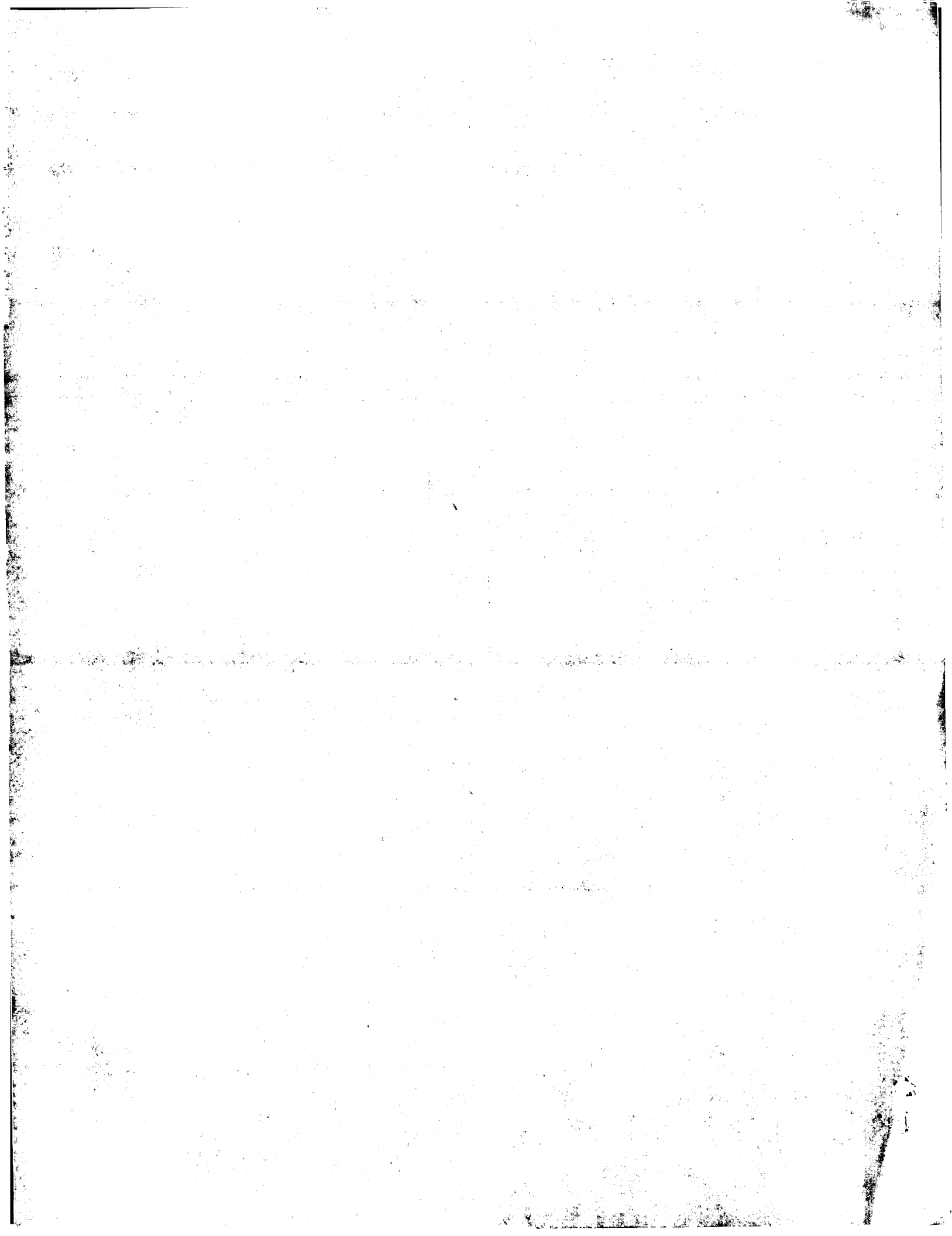
APPLICATION NUMBER: US 60/033,230
FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-984-207-7

Query Match 48.3%; Score 124.5; DB 4; Length 344;
Best Local Similarity 40.6%; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;

QY 3 GSGSGGSGSGSGGQ-----SGSGSGSGSGG-SGSGSGSGSGSGG 43
Db 141 GNGVGANGAKGAGGGGLAELQETIEQLAOLGGGGAGAGGGGCGGAGDGGGAG 200

QY 44 GSGG 47
Db 201 GAGG 204

Search completed: March 13, 2003, 15:06:20
Job time : 3.37192 secs



PR 14-DEC-2000; 2000US-255111P.

[illegible]

PA (CNRS) CENT NAT RECH SCI.
 XX Baubet V, Le Mouellic H, Brulet P;
 XX WPI: 2002-139605/18.
 DR N-PSDB; ABA97889.
 XX New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells
 XX
 PS Claim 6; Page 31; 61pp; English.
 XX
 CC The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEQ
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network.
 XX
 SQ Sequence 477 AA;
 Query Match 100.0%; Score 258; DB 23; Length 477;
 Best Local Similarity 100.0%; Pred. No. 2.4e-16;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50
 Db 239 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 288
 RESULT 4
 ABB08635
 ID ABB08635 standard; Protein; 906 AA.
 XX
 AC ABB08635;
 XX
 DT 03-MAY-2002 (first entry)
 DE
 DE GFP fusion protein Seg5A SEQ ID NO 6.
 XX
 KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
 KW Chemiluminescence Resonance Energy Transfer; CRET;
 KW central nervous system; neural network.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 305 /note= "Encoded by QGT"
 FT Misc-difference 623 /note= "Encoded by CTQ"
 FT Misc-difference 868 /note= "Encoded by GTQ"
 FT
 XX WO200192300-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 01-JUN-2001; 2001WO-EP07057.
 XX
 XX 01-JUN-2000; 2000US-208314P.

PR 09-JUN-2000; 2000US-210526P.
 PR 14-DEC-2000; 2000US-255111P.
 XX (INSP) INST PASTEUR.
 PA (CNRS) CENT NAT RECH SCI.
 XX Baubet V, Le Mouellic H, Brulet P;
 PI WPI: 2002-139605/18.
 DR N-PSDE; ABA97890.
 XX
 PT New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells
 XX
 PS Claim 7; Page 31-32; 61pp; English.
 XX
 CC The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEQ
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network.
 XX
 SQ Sequence 906 AA;
 Query Match 100.0%; Score 258; DB 23; Length 906;
 Best Local Similarity 100.0%; Pred. No. 4.1e-16;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 50
 Db 668 SGGSGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRS 717
 RESULT 5
 ABB08633
 ID ABB08633 standard; Protein; 468 AA.
 XX
 AC ABB08633;
 XX
 DT 03-MAY-2002 (first entry)
 DE
 DE GFP fusion protein G4A SEQ ID NO 4.
 XX
 KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
 KW Chemiluminescence Resonance Energy Transfer; CRET;
 KW central nervous system; neural network.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 401 /note= "Encoded by QGA"
 FT
 XX WO200192300-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 01-JUN-2001; 2001WO-EP07057.
 XX
 XX 01-JUN-2000; 2000US-208314P.

```

PR 09-JUN-2000; 2000US-210526P.
XX 14-DEC-2000; 2000US-255111P.
XX (INSP ) INST PASTEUR.
PA (CNRS ) CENT NAT RECH SCI.
XX Baubet V, Le Mouellie H, Brulet P;
PI WPI; 2002-139605/18.
XX New fusion protein comprising a modified bioluminescent system with a
PT fluorescent molecule covalently linked with a photoprotein, useful for
PT monitoring calcium fluxes or for detecting electrical activity in a
PT group of neural cells
XX Claim 5; Page 31; 61pp; English.
XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy
CC transfer from aequorin to green fluorescent protein by Chemiluminescence
CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
CC covalently linked with a photoprotein. The fusion protein comprises the
CC formula: GFP - LINKER - AEQ
CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
CC polypeptide of 4-63 amino acids.
CC The fusion protein is useful as a bioluminescent Ca+ reporter at the
CC single cell level. The fusion protein, composition or bioluminescent
CC system is useful for monitoring calcium fluxes in real time. This is
CC particularly useful for understanding the development, the plasticity and
CC the functioning of the central nervous system. The fusion protein
CC comprising the bioluminescent system is useful for detecting electrical
CC activity in a group of neural cells, for making it possible to complete
CC the phenotype study of mutants, or for observing the calcium activity in
CC a population of connected cells, for example in a neural network. The
CC present sequence is that of a GFP peptide linker sequence of the
XX invention.
XX Sequence 468 AA;
SQ Query Match 82.2%; Score 212; DB 23; Length 468;
Best Local Similarity 97.6%; Pred. No. 3.5e-12;
Matches 41; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 9 QSGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50
:|||||
DB 238 KSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 279

RESULT 6
ABB08756
ID ABB08756 standard; peptide; 41 AA.
XX ABB08756;
XX ABB08756;
XX 03-MAY-2002 (first entry)
XX GFP peptide linker SEQ ID NO 21.
DE GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
KW Chemiluminescence Resonance Energy Transfer; CRET;
KW central nervous system; neural network.
XX Synthetic.
XX WO200192300-A2.
XX 06-DEC-2001.
XX 01-JUN-2001; 2001WO-EP07057.
XX 01-JUN-2000; 2000US-208314P.
PR 09-JUN-2000; 2000US-210526P.
XX 14-DEC-2000; 2000US-255111P.
XX (INSP ) INST PASTEUR.
PA (CNRS ) CENT NAT RECH SCI.
XX Baubet V, Le Mouellie H, Brulet P;
PI WPI; 2002-139605/18.
XX New fusion protein comprising a modified bioluminescent system with a
PT fluorescent molecule covalently linked with a photoprotein, useful for
PT monitoring calcium fluxes or for detecting electrical activity in a
PT group of neural cells
XX Claim 5; Page 31; 61pp; English.
XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy
CC transfer from aequorin to green fluorescent protein by Chemiluminescence
CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
CC covalently linked with a photoprotein. The fusion protein comprises the
CC formula: GFP - LINKER - AEQ
CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
CC polypeptide of 4-63 amino acids.
CC The fusion protein is useful as a bioluminescent Ca+ reporter at the
CC single cell level. The fusion protein, composition or bioluminescent
CC system is useful for monitoring calcium fluxes in real time. This is
CC particularly useful for understanding the development, the plasticity and
CC the functioning of the central nervous system. The fusion protein
CC comprising the bioluminescent system is useful for detecting electrical
CC activity in a group of neural cells, for making it possible to complete
CC the phenotype study of mutants, or for observing the calcium activity in
CC a population of connected cells, for example in a neural network. The
CC present sequence is that of a GFP peptide linker sequence of the
XX invention.
XX Sequence 41 AA;
SQ Query Match 81.8%; Score 211; DB 23; Length 41;
Best Local Similarity 100.0%; Pred. No. 5.6e-13;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50
|||||
DB 1 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 41

RESULT 7
AAW56163
ID AAW56163 standard; Protein; 738 AA.
XX AAW56163;
XX AAW56163;
XX 28-JUL-1998 (first entry)
XX New DNA sequence isolated from Pinctada fucata.
XX Pinctada fucata; protein; cosmetic.
XX Pinctada fucata.
OS JP10080285-A.
XX JP10080285-A.
XX 31-MAR-1998.
XX 28-MAY-1997; 97JP-0138461.
XX 15-JUL-1996; 96JP-0184459.
XX (MIKI-) MIKIMOTO SEIYAKU KK.
PA WPI; 1998-254410/23.
DR N-PSDB; AAW22683.
XX New cDNA and e.g. vector, host cell and polypeptide - used to
PT produce polypeptide in high yields, which is used in cosmetics
XX
```

Query Match	Score 154;	DB 19;	Length 235;
	59.78;		

[illegible]

Result No.	Score		Query		Length	DB	ID	Description
	Score	Match	Match	Match				
1	147.5	57.2	532	10	US-09-833-790-428	Sequence 428, Appl		
2	147	57.0	522	10	US-09-848-990-22	Sequence 22, Appl		
3	147	57.0	201	10	US-09-760-364-14	Sequence 14, Appl		
4	147	57.0	357	10	US-09-864-761-35807	Sequence 35807, A		
5	146	56.6	200	9	US-10-160-354-4	Sequence 4, Appl		
6	146	56.6	200	9	US-09-990-940-21	Sequence 21, Appl		
7	146	56.6	200	9	US-10-026-021-8	Sequence 8, Appl		
8	146	56.6	200	9	US-10-161-165-3	Sequence 3, Appl		
9	146	56.6	200	9	US-10-160-663-3	Sequence 3, Appl		
10	146	56.6	200	9	US-10-071-838-15	Sequence 15, Appl		
11	146	56.6	200	9	US-10-094-417-25	Sequence 25, Appl		
12	146	56.6	200	10	US-09-798-584-18	Sequence 18, Appl		
13	146	56.6	200	10	US-09-967-624-19	Sequence 19, Appl		
14	146	56.6	200	10	US-09-998-667-18	Sequence 18, Appl		
15	144	55.8	40	9	US-10-081-400-1	Sequence 1, Appl		
16	139.5	54.1	618	10	US-09-935-300-1381	Sequence 1381, Ap		
17	138	53.5	484	9	US-09-820-843A-19	Sequence 19, Appl		
18	137	53.1	283	10	US-09-864-761-36720	Sequence 36720, A		
19	136.5	52.9	40	9	US-10-005-438-9	Sequence 9, Appl		


```

; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-990-940-21

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY   2  GSGSGGQGSGGGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |
Db    103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 7
US-10-026-021-8
; Sequence 8, Application US/10026021
; Publication No. US20030027756A1
; GENERAL INFORMATION:
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Demo, Susan
; APPLICANT: Jenkins, Yonchu
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: SAK: Modulation of Cellular Proliferation for
; TITLE OF INVENTION: Treatment of Cancer
; FILE REFERENCE: 021044-001210US
; CURRENT APPLICATION NUMBER: US/10/026,021
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/309,632
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-026-021-8

Query Match          56.6%; Score 146; DB 9; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY   2  GSGSGGQGSGGGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 47
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |
Db    103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 8
US-10-161-165-3
; Sequence 3, Application US/10161165
; Publication No. US20030027763A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Mark
; APPLICANT: Holland, Sacha
; APPLICANT: Rossi, Alex
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CD43: Modulators of Mast Cell Degranulation
; FILE REFERENCE: 021044-001010US
; CURRENT APPLICATION NUMBER: US/10/161,165

```


[illegible]

RESULT 12

```

US-09-798-584-18
; Sequence 18, Application US/09798584
; Patent No. US20020102676A1
; GENERAL INFORMATION:
; APPLICANT: Mu, David
; APPLICANT: Powers, Scott
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: KCNB: A No. US20020102676A1e1 Potassium Channel Protein
; FILE REFERENCE: 018781-004010US
; CURRENT APPLICATION NUMBER: US/09/798,584
; CURRENT FILING DATE: 2001-03-03
; PRIOR APPLICATION NUMBER: US 60/186,951
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly-Gly linker
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
; US-09-798-584-18

```

Query Match	56.6%	Score 146;	DB 10;	Length 200;
Best Local Similarity	56.5%	Pred. NO. 1.5e-05;		
Matches 26; Conservative	0;	Mismatches 20;	Indels	

Qy 2 GGSGGGQSGGSGCGQSGGSGGGQSGGSGSGGSGGSG 47
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 GGG 148

RESULT 13

```

US-09-967-624-19
; Sequence 19, Application US/09967624
; Patent No. US20020142325A1
; GENERAL INFORMATION:
; APPLICANT: Liao, X. Charlene
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Li, Congfen
; APPLICANT: Zhao, Haoran
; APPLICANT: Wu, Jun
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: PAK2: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-000700US
; CURRENT APPLICATION NUMBER: US/09/967,624
; PRIORITY FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/380,647
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly G1
; OTHER INFORMATION: flexible linker

```

```

; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present

```

US-09-967-624-19

Query Match 56.6%; Score 146; DB 10; Length 200;
Best Local Similarity 56.5%; Pred. No. 1.5e-05;
Matches 26; Conservative 0; Mismatches 20; Indels

Qy 2 GGSGGGGSGSGSGGSGSGSGSGSGSGSGSGSG 47
 || | | | | | | | | | | | | | | | | |
Dz 103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148
 || | | | | | | | | | | | | | | | | |

RESULT 14

```

US-09-998-667-18
; Sequence 18, Application US/09998667
; Patent No. US2020146747A1
; GENERAL INFORMATION:
; APPLICANT: Masuda, Esteban
; APPLICANT: Liao, X. Charlene
; APPLICANT: Zhao, Haoran
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: TRAC1: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-000600US
; CURRENT APPLICATION NUMBER: US/09/998,667
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: US 60/282,432
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: flexible
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or ab
US-09-998-667-18

```

Query Match	56.6%	Score 146;	DB 10;	Length 200;
Best Local Similarity	56.5%;	Pred. No. 1.5e-05;		
Matches 26;	Conservative	0;	Mismatches 20;	Indels 0;
				Gaps 0;

Qy 2 GSGSGGQSGSGSGGQSGSGSGGQSGSGSGGQSGG 47
 || || || || || || || || || || || || || || || ||
pb 103 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 148

RESULT 15

```

RES-13
US-10-081-400-1
: Sequence 1, Application US/10081400
: Patent No. US2002015598A1
: GENERAL INFORMATION:
: APPLICANT: Young, Michael
: APPLICANT: Meade, Harry
: APPLICANT: Krane, Ian
: TITLE OF INVENTION: ERYTHROPOIETIN ANALOGY-HUMAN SERUM ALBUMIN FUSION
: FILE REFERENCE: 10275/041001
: CURRENT APPLICATION NUMBER: US/10/081,400
: CURRENT FILING DATE: 2002-02-20
: PRIOR APPLICATION NUMBER: US 09/333,213
: PRIOR FILING DATE: 1999-06-15
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 40
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:

```

Search completed: March 13, 2003, 15:12:36
Job time : 3.08729 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	170	65.9	302	2	C84470	hypothetical prote
2	168	65.1	592	2	E82759	endo-1,4-beta-gluc
3	161	62.4	115	2	D61615	sericin MG-1 - gre
4	159.5	61.8	207	2	T07381	glycine-rich prote
5	154	59.7	641	1	Q0BE31	nuclear antigen BB
6	153	59.3	1218	2	E84537	hypothetical prote
7	148.5	57.6	171	2	H84709	probable glycine-r
8	146	56.6	291	1	S31415	glycine-rich prote
9	146	56.6	622	2	I37984	keratin 9, type I,
10	145	56.2	1226	2	T24045	hypothetical prote
11	144.5	56.0	221	2	T04592	glycine-rich cell
12	144	55.8	183	2	PN0109	keratin-like prote
13	144	55.8	569	1	KRMSE1	keratin, 59K type
14	143.5	55.6	208	2	T46896	merozoite surface
15	143.5	55.6	593	1	KRRU0	keratin 10, type I
16	143	55.4	465	1	S01820	glycine-rich cell
17	143	55.4	526	1	KRBOV1	keratin, 54K type
18	142.5	55.2	165	1	KNRZG1	glycine-rich cell
19	142.5	55.2	166	1	KRB02B	keratin, 68K type
20	142.5	55.2	570	2	S07330	keratin, epidermal
21	142	55.0	481	2	A35628	loricatin - mouse
22	142	55.0	581	1	KRMS2	keratin, type II c
23	141	54.7	396	2	T49109	glycine-rich prote
24	139	53.9	183	1	KNRZG2	glycine-rich cell
25	138.5	53.7	271	2	S34666	glycine-rich prote
26	138.5	53.7	434	1	Z3BPIK	coat protein A - p
27	138	53.5	484	2	G70846	hypothetical glyci
28	137.5	53.3	274	2	T25404	hypothetical prote
29	137	53.1	320	2	T09555	fibrillarin - Arab

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	161.5	62.6	450	1	SWP1_ENCCU	O9XZV1 encephalito
2	161	62.4	115	1	SER1_CALME	O96614 galleria me
3	154	59.7	641	1	EBN1_EBV	O93211 epstein-bar
4	147.5	57.2	532	1	ZIC2_HUMAN	O95409 homo sapien
5	146	56.6	622	1	K1C1_HUMAN	P35527 homo sapien
6	144	55.8	569	1	K1CJ_MOUSE	O92535 mus musculu
7	143.5	55.6	593	1	K1CJ_HUMAN	P13645 homo sapien
8	143	55.4	465	1	GRP2_BOVIN	P10496 phaseolus v
9	143	55.4	526	1	K1CJ_BOVIN	O96394 bos taurus
10	143	55.4	677	1	SP87_DICDI	P54643 dictyosteli
11	142.5	55.2	165	1	GRP1_ORYSA	P25074 oryza sativ
12	142.5	55.2	166	1	K2C5_BOVIN	O94262 bos taurus
13	142	55.0	481	1	L2C1_MOUSE	P18165 mus musculu
14	141	54.7	627	1	K2C1_MOUSE	O94104 mus musculu
15	139	53.9	183	1	GRP2_ORYSA	P29834 oryza sativ
16	138.5	53.7	434	1	COAA_BPICE	O93663 bacterioph
17	138	53.5	1250	1	TP3A_DROME	O90698 drosophila
18	137	53.1	384	1	GRP1_PETHY	O95789 petunia hyb
19	136	52.7	333	1	SIX3_MOUSE	O62233 mus musculu
20	135	52.3	214	1	GRP2_NICYS	P27484 nicotiana s
21	135	52.3	1901	1	YZ08_MYCTU	O53553 mycobacteri
22	133.5	51.7	316	1	LOR1_HUMAN	P23490 homo sapien
23	133	51.6	252	1	GRP1_PHAVU	P10495 phaseolus v
24	132	51.2	386	1	RB87_DROME	O48810 drosophila
25	131.5	51.0	463	1	YA68_MYCTU	O53416 mycobacteri
26	131.5	51.0	491	1	YK98_MYCTU	O10707 mycobacteri
27	131	50.8	1380	1	DDX9_MOUSE	O70133 mus musculu
28	131	50.8	3178	1	YS89_CAEEL	O09624 caenorhabdi
29	130.5	50.6	434	1	COAA_BP122	P15415 bacterioph
30	130.5	50.6	643	1	K2C1_HUMAN	P04264 homo sapien
31	130	50.4	440	1	FXGA_CHICK	O98937 gallus gall
32	130	50.4	763	1	GLH1_CAEEL	P34689 caenorhabdi
33	130	50.4	1726	1	MSP1_PLAFC	O94934 plasmodium

FT	CONFLICT	24	S -> F (IN REF. 2).
FT	CONFLICT	28	S -> F (IN REF. 2).
FT	CONFLICT	38	E -> L (IN REF. 2).
FT	CONFLICT	41	E -> G (IN REF. 2).
FT	CONFLICT	104	AG -> GS (IN REF. 2).
FT	CONFLICT	110	MISSING (IN REF. 2).
FT	CONFLICT	121	SY -> GC (IN REF. 2).
FT	CONFLICT	137	SY -> G (IN REF. 2).
FT	CONFLICT	148	Q -> R (IN REF. 2).
FT	CONFLICT	178	WYKHHGNSO -> VVREAROLKP (IN REF. 2).
FT	CONFLICT	263	KSDLEM -> QSVLEL (IN REF. 2).
FT	CONFLICT	284	H -> L (IN REF. 2).
FT	CONFLICT	353	E -> A (IN REF. 2).
FT	CONFLICT	394	GGRYCV -> VESLLR (IN REF. 2).
FT	CONFLICT	508	GSGHGS -> CGSRRG (IN REF. 2).
FT	CONFLICT	523	S -> G (IN REF. 2).
FT	CONFLICT	531	H -> R (IN REF. 2).
FT	CONFLICT	534	S -> G (IN REF. 2).
FT	CONFLICT	543	S -> G (IN REF. 2).
FT	CONFLICT	547	GQ -> RR (IN REF. 2).
FT	CONFLICT	555	KS -> SGT (IN REF. 2).
ST	SEQUENCE	569 AA;	57711 MW; BEC59D4D8FFE484D CRC64;
ST	SEQUENCE	55.88;	Score 144; DB 1; Length 569;
ST	Best Local Similarity	60.48;	Pred. No. 0.00018;
ST	Matches 29; Conservative	0; Mismatches 17; Indels 1;	Gaps 1;
QY	2	GGSSGGSGGSGGGSGGSGQS--GGSSGGSGGSGGSGGSGG 47	
DB	503	GGSSGGSGGSGGGGSGGSSGACGHGSSGGGSGGSGGSGG 550	
RESULT 7			
KICJ_HUMAN	STANDARD;	PRT;	593 AA.
ID	KICJ_HUMAN		
AC	P13645;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	keratin, type I cytoskeletal 10 (Cyto keratin 10) (K10) (CK 10).		
GN	KT10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89125611; PubMed=2464696;		
RA	Rieger M., Franke W.W.;		
RT	"Identification of an orthologous mammalian cytokeratin gene. High degree of intron sequence conservation during evolution of human cytokeratin 10.";		
RL	J. Mol. Biol. 204:841-856(1988).		
RN	[2]		
RP	SEQUENCE OF 130-593 FROM N.A.		
RX	MEDLINE=88122104; PubMed=2448602;		
RA	Darmon M.Y., Senat A., Darmon M.C., Vasquez M.;		
RT	"Sequence of a cDNA encoding human keratin No 10 selected according to structural homologies of keratins and their tissue-specific expression.";		
RL	J. Mol. Biol. Rep. 12:277-283(1987).		
RN	[3]		
RP	SEQUENCE OF 197-593 FROM N.A.		
RX	MEDLINE=92339897; PubMed=1378806;		
RA	Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P., Kisselev L.L.;		
RT	"Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo structural rearrangements within repeats.";		
RL	Gene 116:245-251(1992).		
RN	[4]		
RP	SEQUENCE OF 180-184 AND 577-589.		
RC	TISSUE=Keratinocytes;		
RX	MEDLINE=93162043; PubMed=12866667;		

Rasmussen H.H., van Damme J., Puype M., Gesser B., Cellis J.E., Vandekerckhove J.;

"Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes.";

Electrophoresis 13:960-969(1992).

[5]

VARIANT EHK HIS-156.

MEDLIN=92386600; PubMed=1381287;

Cheng J., Syder A.J., Yu Q.-C., Letai A., Paller A.S., Fuchs E.;

"The genetic basis of epidermolytic hyperkeratosis: a disorder of differentiation-specific epidermal keratin genes.";

Cell 70:811-819(1992).

[6]

VARIANTS.

MEDLIN=92141228; PubMed=1371013;

Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;

"Extensive size polymorphism of the human keratin 10 chain resides in the C-terminal V2 subdomain due to variable numbers and sizes of glycine loops.";

Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).

[7]

VARIANTS EHK HIS-156 AND SER-161.

MEDLIN=92376531; PubMed=1380725;

Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A., Greenhaigh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;

"Mutations in the rod domains of keratins 1 and 10 in epidermolytic hyperkeratosis.";

Science 257:1128-1130(1992).

[8]

VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.

MEDLIN=94136477; PubMed=7508181;

Chihev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L., Compton J.G., Bale S.J.;

"Preferential sites in keratin 10 that are mutated in epidermolytic hyperkeratosis.";

Am. J. Hum. Genet. 54:179-190(1994).

[9]

VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.

MEDLIN=94216497; PubMed=7512983;

Syde A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;

"Genetic mutations in the K1 and K10 genes of patients with epidermolytic hyperkeratosis. Correlation between location and disease severity.";

J. Clin. Invest. 93:1533-1542(1994).

[10]

VARIANT EHK ASN-160.

MEDLIN=94117868; PubMed=75071150;

Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;

"Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene sequencing.";

J. Invest. Dermatol. 102:13-16(1994).

[11]

VARIANTS EHK PRO-156 AND SER-156.

MEDLINE=94117870; PubMed=75071152;

McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R., Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G., Morley S.M.;

"Mutations in the rod 1A domain of keratins 1 and 10 in bullous congenital ichthyosiform erythroderma (BCIE).";

J. Invest. Dermatol. 102:24-30(1994).

[12]

VARIANT EHK THR-150.

MEDLINE=95059228; PubMed=7526210;

Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G., Fuchs E.;

"Genetic and clinical mosaicism in a type of epidermal nevus.";

New Engl. J. Med. 331:1408-1415(1994).

[13]

VARIANT AEI THR-446.

MEDLINE=99072665; PubMed=9856845;

Suga Y., Duncan K.O., Heald P.W., Roop D.R.;

"A novel helix termination mutation in keratin 10 in annular epidermolytic ichthyosis, a variant of bullous congenital epidermolytic ichthyosis.";

RESULT 10
SP87_DICDI

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:59 ; Search time 4.93359 Seconds
(without alignments)
2088.208 Million cell updates/sec

Title: US-09-863-901-22

Perfect score: 258

Sequence: 1 SCGSGGSGGSGGSGGSGG.....SGSGGSGGSGGSGGSGGLRS 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	170	65.9	302	10 Q9SL09	Q9SL09 arabidopsis
2	170	65.9	738	5 O02402	O02402 pinctada fu
3	169	65.5	125	5 Q964C1	Q964C1 encephalito
4	168	65.1	592	16 Q9PF60	Q9PF60 xylella fas
5	166	64.3	102	5 Q964C0	Q964C0 encephalito
6	163	63.2	1422	10 Q9ZUR3	Q9ZUR3 arabidopsis
7	163	63.2	1745	10 Q9MAK1	Q9MAK1 arabidopsis
8	161.5	62.6	100	5 Q964C3	Q964C3 encephalito
9	161.5	62.6	117	5 Q964C2	Q964C2 encephalito
10	161.5	62.6	121	5 Q964C5	Q964C5 encephalito
11	161.5	62.6	138	5 Q964C4	Q964C4 encephalito
12	159.5	61.8	207	10 Q43522	Q43522 lycopersico
13	154	59.7	786	6 O18740	O18740 canis famil
14	153	59.3	1218	10 Q9XH2	Q9XH2 arabidopsis
15	150	58.1	251	3 Q9UCV7	Q9UCV7 yarrowia li
16	149	57.8	157	5 Q9GQX8	Q9GQX8 plasmodium

ALIGNMENTS

RESULT 1

ID	Q9SL09	PRELIMINARY;	PRT;	302 AA.
AC	Q9SL09;			
DT	01-MAY-2000 (Tremblrel. 13, Created)			
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	At2g05580 protein.			
GN	AT2G05580.			
OS	Arabidopsis thaliana (Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid3 II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RX	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,			
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,			
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.;			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RT	thaliana.";			
RL	Nature 402:761-768(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Lin X.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC006220; AAC24649.1; .			
DR	InterPro; IPR002952; Eggshell.			
DR	InterPro; IPR000817; PRION.			
DR	PRINTS; PR01228; EGGSHLL.			
DR	PRINTS; PR00341; PRION.			
SQ	SEQUENCE 302 AA; 26008 MW; E72A73C55825891E CRC64;			

17	148.5	57.6	171	10	O04339	O04339 arabidopsis
18	147.5	57.2	87	10	Q9LRI5	Q9LRI5 pisum sativ
19	146	56.6	291	10	Q39337	Q39337 brassica na
20	146	56.6	370	5	Q9BPK8	Q9BPK8 plasmodium
21	146	56.6	370	5	Q9BPK6	Q9BPK6 plasmodium
22	145.5	56.4	118	10	Q949R3	Q949R3 arabidopsis
23	145	56.2	277	5	Q9GQP0	Q9GQP0 plasmodium
24	145	56.2	302	5	Q9GQL9	Q9GQL9 plasmodium
25	145	56.2	302	5	Q9COL8	Q9COL8 plasmodium
26	145	56.2	317	5	Q9BH54	Q9BH54 plasmodium
27	144.5	56.0	221	10	O65514	O65514 arabidopsis
28	144	55.8	561	11	Q9CXH6	Q9CXH6 mus musculu
29	143.5	55.6	208	5	Q25949	Q25949 plasmodium
30	143.5	55.6	718	12	Q91TR1	Q91TR1 tupiaia herp
31	143	55.4	588	12	Q9QEK6	Q9QEK6 cynomolgus
32	142	55.0	486	11	Q8R019	Q8R019 mus musculu
33	142	55.0	698	12	Q8QKX8	Q8QKX8 ectocarpus
34	141.5	54.8	619	12	Q9IP09	Q9IP09 cynomolgus
35	141	54.7	139	5	Q9GQX9	Q9GQX9 plasmodium
36	141	54.7	208	5	Q9U0A0	Q9U0A0 plasmodium
37	141	54.7	396	10	O65450	O65450 arabidopsis
38	140	54.3	610	5	Q9V5V8	Q9V5V8 drosophila
39	139.5	54.1	637	11	Q9D2K8	Q9D2K8 mus musculu
40	139	53.9	185	10	Q948R3	Q948R3 oryza sativ
41	138.5	53.7	271	10	O08529	O08529 nicotiana t
42	138.5	53.7	588	12	Q9IP08	Q9IP08 cynomolgus
43	138	53.5	164	5	Q9BJQ5	Q9BJQ5 plasmodium
44	138	53.5	484	16	O53394	O53394 mycobacteri
45	137.5	53.3	175	10	Q9LSN6	Q9LSN6 arabidopsis

Xiao L., Li L., Visvesvara G.S., Moura H., Didier E.S., Lal A.A.;
 "Genotyping Encephalitozoon cuniculi by Multilocus Analyses of Genes
 with Repetitive Sequences";
 J. Clin. Microbiol. 39:2248-2253(2001).
 DR EMBL: AF40012; AAK63052.1; -;
 FT NON_TER 1
 FT NON_TER 102
 FT NON_TER 102
 FT NON_SEQUENCE 102 AA: 8415 MW: C5AB107795B2EC6D CRC64:
 SO SEQUENCE

OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta, eudicotyledons: core eudicots: Rosidae;
OC eucosids II: Brassicales; Brassicaceae; Arabidopsis.

```

00  euosios ii, brassicae, brassicae, brassicae;
0X  NCBI_TaxID=3702;

```

RP SEQUENCE FROM N.A.
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E. Conn L.,

RA Conway A., Gonzales A., Hansen N., Howay B., Koo R., Lam B., Lee C.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,

RA Ecker J.R.;

RT
I. " ;

DR EMBL; AC007915; AAF69170.1; -.

DR Interpro; IPR000087; Collagen.
DR Interpro: TPR003653: S1MO protease.

DR P1am; PF02902; peptidase_C48; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.

TO: DIRECTOR, FBI (100-388610)
FROM: SAC, NEW YORK (100-100000)
SUBJECT: MURDER OF MARTIN LUTHER KING, JR.
RE: NEW YORK TELETYPE TO BUREAU, APRIL 4, 1968.

Best Local Similarity 61.7%; Pred. No. 5.4e-07;

A7

dh 426 SGGDGGGPPSGGGGPPSGGGGPPSGGGGPPSGGGGPPNG 472

RESOLUT
Q964C3

AC Q964C3:

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE Spore wall protein-1 (Fragment).

OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI TaxID=6035.

RN
[I]
RP
SEQUENCE FROM N. A.

KC STRAIN=STRAIN 11;
 RX MEDLINE=21270266; PubMed=11376065;

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
8

[illegible]

FT NON_TER 1 1
SQ SEQUENCE 251 AA; 21773 MW; 4BC3E20ECF45C825 CRC64;
Query Match 58.1%; Score 150; DB 3; Length 251;
Best Local Similarity 66.7%; Pred. No. 1.2e-06;
Matches 34; Conservative 0; Mismatches 13; Indels 4; Gaps 2;
QY 1 SGGSGGGQGGGGGGGG--GGSGGGGGGGGGGG--GGSGGGGG 47
Db 73 SGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123

Search completed: March 13, 2003, 15:05:07
Job time : 6.93359 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 13, 2003, 14:59:19 ; Search time 58.833 Seconds
(without alignments)
1080.355 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGEELTGVPILVELDG.....LGFWTMPDACEKLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	100.0	477	23 ABB08634	GFP fusion protein
2	2560	100.0	906	23 ABB08635	GFP fusion protein
3	2496.5	97.6	468	23 ABB08633	GFP fusion protein
4	2395.5	93.6	450	23 ABB08632	GFP fusion protein
5	2344	91.6	441	23 ABB08631	GFP fusion protein
6	2287.5	89.4	432	23 ABB08630	GFP fusion protein
7	1353	52.9	655	23 ABB75672	Protein related to
8	1295	50.6	719	19 AAW85012	Smad2-green flour
9	1295	50.6	719	21 AAY70779	EGFP-SMAD2 fusion
10	1293.5	50.5	356	23 ABB08620	Yeast polypeptide

11	1292.5	50.5	631	19 AAW85006	Erk1-green floures
12	1292.5	50.5	631	21 AAY70778	EGFP-Erk1 fusion p
13	1291.5	50.4	1386	21 AAY85573	Hs-UNC-53/3 fragm
14	1290.5	50.4	890	21 AAB22938	GFP-HSC70 fusion p
15	1288.5	50.3	294	21 AAB22860	GFP-DEVD-annexin I
16	1288.5	50.3	294	21 AAY79638	Caspase-3 biosenso
17	1286	50.2	1070	18 AAW17789	Green fluorescent
18	1283.5	50.1	633	19 AAW85009	Jnk1-green floures
19	1283	50.1	459	21 AAB22936	GFP-HSP27 fusion p
20	1278.5	49.9	635	19 AAW85034	Green fluorescent
21	1278.5	49.9	635	21 AAY70781	EGFP-VASP fusion p
22	1277	49.9	501	18 AAW31879	GFP variants S65T
23	1276.5	49.9	1171	19 AAW85037	Green fluorescent
24	1274	49.8	250	22 AAG66492	Green fluorescence
25	1272	49.7	558	19 AAW48662	GR fusion protein
26	1271.5	49.7	1090	21 AAY85576	Hs-UNC-53/1 fragm
27	1271	49.6	397	20 AAY42176	EGFP/DRM fusion pr
28	1271	49.6	429	20 AAY42175	EGFP/DRM fusion pr
29	1271	49.6	501	18 AAW31877	GFP variants S65C
30	1271	49.6	501	18 AAW31878	GFP variants S65C
31	1271	49.6	514	18 AAW31876	GFP variants S65C
32	1270	49.6	238	18 AAW22101	Aequorea victoria
33	1270	49.6	238	20 AAW96328	Humanised green fl
34	1270	49.6	247	20 AAW96329	Humanised green fl
35	1269.5	49.6	359	20 AAY42179	EGFP/DRM fusion pr
36	1269.5	49.6	359	20 AAY42180	EGFP/DRM fusion pr
37	1269.5	49.6	391	20 AAY42178	EGFP/DRM fusion pr
38	1269.5	49.6	997	19 AAW85032	Green fluorescent
39	1269	49.6	607	19 AAW85010	p38-green flouresc
40	1269	49.6	783	21 AAB22939	GFP-HSFl fusion pr
41	1268	49.5	248	23 AAG68319	Jellyfish green fl
42	1268	49.5	403	20 AAY42177	EGFP/DRM fusion pr
43	1268	49.5	842	19 AAW85008	Grk5-green floures
44	1267	49.5	605	19 AAW85007	Erk2-green floures
45	1267	49.5	727	19 AAW85041	Green fluorescent

ALIGNMENTS

... RESULT 1

ABB08634

ID ABB08634 standard; Protein: 477 AA.

XX ABB08634;

XX 03-MAY-2002 (first entry)

DE GFP fusion protein G5A SEQ ID NO 5.

XX GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
XX Chemiluminescence Resonance Energy Transfer; CREt;
XX central nervous system; neural network.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 445 /note= "Encoded by AQT"

XX WO200192300-A2.

XX 06-DEC-2001.

XX 01-JUN-2001; 2001WO-EP07057.

XX 01-JUN-2000; 2000US-208314P.

PR 09-JUN-2000; 2000US-210526P.

PR 14-DEC-2000; 2000US-255111P.

XX (INSP ; INST PASTEUR.

PA (CNRS ; CENT NAT RECH SCI.

XX

[illegible]

CC	transfer from aequorin to green fluorescent protein by Chemiluminescence
CC	Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
CC	covalently linked with a photoprotein. The fusion protein comprises the
CC	formula: GFP - LINKER - AEO
CC	GFP = green fluorescent protein; AEO = aequorin; and LINKER = a
CC	polypeptide of 4-63 amino acids.
CC	The fusion protein is useful as a bioluminescent Ca ²⁺ reporter at the
CC	single cell level. The fusion protein, composition or bioluminescent
CC	system is useful for monitoring calcium fluxes in real time. This is
CC	particularly useful for understanding the development, the plasticity and
CC	the functioning of the central nervous system. The fusion protein
CC	comprising the bioluminescent system is useful for detecting electrical
CC	activity in a group of neural cells, for making it possible to complete
CC	the phenotype study of mutants, or for observing the calcium activity in
CC	a population of connected cells, for example in a neural network.
XX	
SQ	Sequence 468 AA;
	Query Match 97.6%; Score 2498.5; DB 23; Length 468;
	Best Local Similarity 98.1%; Pred. No. 2.le-202;
	Matches 468; Conservative 0; Mismatches 0; Indels 9; Gaps 1;
QY	1 MSKGEELFTGVVPLVELDGVNHHKFSVSSEGEGDATYGKLTLKFICTTCKLPVPWPTL 60
DB	1 MSKGEELFTGVVPLVELDGVNHHKFSVSSEGEGDATYGKLTLKFICTTCKLPVPWPTL 60
QY	61 VTLLTYGVOCESRYPDHMKQHDFKFSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
DB	61 VTLLTYGVOCESRYPDHMKQHDFKFSAMPEGYVOERTIFFKDDGNKYKTRAEVKFEGDTLV 120
QY	121 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKKNGIKANPKIRHNIEDGSVOLAD 180
DB	121 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKKNGIKANPKIRHNIEDGSVOLAD 180
QY	181 HYQONTPIGDGPVLLPDNHXYLSTQSALSKDPNEKRHDHWLLEFVTAAGITHGMDELYKSG 240
DB	181 HYQONTPIGDGPVLLPDNHXYLSTQSALSKDPNEKRHDHWLLEFVTAAGITHGMDELYK-- 238
QY	241 GSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRSVKLTSDFDNPRW 300
DB	239 -----S GSGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGLRSVKLTSDFDNPRW 291
QY	301 IGRHKHMNFLDVNHHGKISLDEMYYKASDIVINNLGATPEQAKRHKDAVEAFFGGAGMK 360
DB	292 IGRHKHMNFLDVNHHGKISLDEMYYKASDIVINNLGATPEQAKRHKDAVEAFFGGAGMK 351
QY	361 YGVETDPMPAYIEGWKKLATDELEYAKNEPFLIRIGWDALFDIVDKONGAITLDEWKAY 420
DB	352 YGVETDPMPAYIEGWKKLATDELEYAKNEPFLIRIGWDALFDIVDKONGAITLDEWKAY 411
QY	421 TKAAGITQSSDCBETFRVCDIDESGQLDVDEMTRQHLGFWMYTMPACEKLYGGAVP 477
DB	412 TKAAGITQSSDCBETFRVCDIDESGQLDVDEMTRQHLGFWMYTMPACEKLYGGAVP 468
RESULT 4	
ABB08632	
ID	ABB08632 standard; Protein; 450 AA.
XX	
AC	ABB08632;
XX	
XX	03-MAY-2002 (first entry)
DT	
XX	GFP fusion protein G2A SEQ ID NO 3.
DE	
XX	
KW	GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
KW	Chemiluminescence Resonance Energy Transfer; CRET;
KW	central nervous system; neural transfer.
XX	
OS	Synthetic.
XX	
PX	WO200192300-A2.
XX	

PD 06-DEC-2001.

XX 01-JUN-2001; 2001WO-EP07057.

XX 01-JUN-2000; 2000US-208314P.

XX 09-JUN-2000; 2000US-210526P.

XX 14-DEC-2000; 2000US-255111P.

XX (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

XX Baubet V, Le Mouellie H, Brulet P;

XX WPI; 2002-139605/18.

DR N-PSDB; ABA97887.

XX New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for

PT monitoring calcium fluxes or for detecting electrical activity in a

PT group of neural cells -

XX Synthetic.

XX WO2001923300-A2.

XX 06-DEC-2001.

XX 01-JUN-2001; 2001WO-EP07057.

XX 01-JUN-2000; 2000US-208314P.

XX 09-JUN-2000; 2000US-210526P.

XX 14-DEC-2000; 2000US-255111P.

XX (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

XX Baubet V, Le Mouellie H, Brulet P;

XX WPI; 2002-139605/18.

DR N-PSDB; ABA97886.

XX New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for

PT monitoring calcium fluxes or for detecting electrical activity in a

PT group of neural cells -

XX Claim 4; Page 31; 61pp; English.

XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy

CC transfer from aequorin to green fluorescent protein by chemiluminescence

CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule

CC covalently linked with a photoprotein. The fusion protein comprises the

CC formula: GFP - LINKER - AEG

CC GFP = green fluorescent protein; AEG = aequorin; and LINKER = a

CC polypeptide of 4-63 amino acids.

CC The fusion protein is useful as a bioluminescent Ca+ reporter at the

CC single cell level. The fusion protein, composition or bioluminescent

CC system is useful for monitoring calcium fluxes in real time. This is

CC particularly useful for understanding the development, the plasticity and

CC the functioning of the central nervous system. The fusion protein

CC comprising the bioluminescent system is useful for detecting electrical

CC activity in a group of neural cells, for making it possible to complete

CC the phenotype study of mutants, or for observing the calcium activity in

CC a population of connected cells, for example in a neural network.

XX SQ Sequence 450 AA;

Query Match 93.6%; Score 2395.5; DB 23; Length 450;

Best Local Similarity 94.3%; Pred. No. 9.8e-194;

Matches 450; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

QY 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTIKFKICTTGKLPVPWPTL 60

DB 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTIKFKICTTGKLPVPWPTL 60

QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 120

DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 120

QY 121 NRIELKGDIFDEGDNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180

DB 121 NRIELKGDIFDEGDNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDMHVLLEFFVTAAGTHGMDELYKSG 240

DB 181 HYQNTPTIGDGPVLLPDNHYLSTQALSCKDPNEKRDMHVLLEFFVTAAGTHGMDELYK-- 238

QY 241 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRVKLTSDFDNPRW 300

DB 239 -----SGGSGGSGGSGGSGGSGGSGGSGGSLRVKLTSDFDNPRW 273

QY 301 IGRHKHMFNFDVNHNGKISIDEMVYKASDIVINNLGATPEQAKRHKDAVEAFEGGAGMK 360

DB 274 IGRHKHMFNFDVNHNGKISIDEMVYKASDIVINNLGATPEQAKRHKDAVEAFEGGAGMK 333

QY 361 XGVETDWPAYIEGWKKLATDELEYAKNEPTLIRIWGDALEFDIVDKDQNGAITLDEWKAY 420

DB 334 XGVETDWPAYIEGWKKLATDELEYAKNEPTLIRIWGDALEFDIVDKDQNGAITLDEWKAY 393

QY 421 TKAAGIIOSSEDCETFRVCDIDESGQLDVEDMTQRHLGFWYTMDPACEKLYGGAVP 477

DB 394 TKAAGIIOSSEDCETFRVCDIDESGQLDVEDMTQRHLGFWYTMDPACEKLYGGAVP 450

RESULT 5

ABB08631

ID ABB08631 standard; Protein; 441 AA.

XX AC ABB08631;

XX DT 03-MAY-2002 (first entry)

XX DE GFP fusion protein G1A SEQ ID NO 2.

XX KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;

XX KW Chemiluminescence Resonance Energy Transfer; CRET;

XX KW central nervous system; neural network.

XX OS Synthetic.

XX PN WO2001923300-A2.

XX PD 06-DEC-2001.

XX PF 01-JUN-2001; 2001WO-EP07057.

XX PR 01-JUN-2000; 2000US-208314P.

XX PR 09-JUN-2000; 2000US-210526P.

XX PR 14-DEC-2000; 2000US-255111P.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CENT NAT RECH SCI.

XX PI Baubet V, Le Mouellie H, Brulet P;

XX WPI; 2002-139605/18.

DR N-PSDB; ABA97886.

XX New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for

PT monitoring calcium fluxes or for detecting electrical activity in a

PT group of neural cells -

XX Claim 3; Page 30-31; 61pp; English.

XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy

CC transfer from aequorin to green fluorescent protein by chemiluminescence

CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule

CC covalently linked with a photoprotein. The fusion protein comprises the

CC formula: GFP - LINKER - AEG

CC GFP = green fluorescent protein; AEG = aequorin; and LINKER = a

CC polypeptide of 4-63 amino acids.

CC The fusion protein is useful as a bioluminescent Ca+ reporter at the

CC single cell level. The fusion protein, composition or bioluminescent

CC system is useful for monitoring calcium fluxes in real time. This is

CC particularly useful for understanding the development, the plasticity and

CC the functioning of the central nervous system. The fusion protein

CC comprising the bioluminescent system is useful for detecting electrical

CC activity in a group of neural cells, for making it possible to complete

CC the phenotype study of mutants, or for observing the calcium activity in

CC a population of connected cells, for example in a neural network.

XX SQ Sequence 441 AA;

Query Match 91.6%; Score 2344; DB 23; Length 441;

Best Local Similarity 92.5%; Pred. No. 2.1e-189;

Matches 441; Conservative 0; Mismatches 0; Indels 36; Gaps 1;

QY 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTIKFKICTTGKLPVPWPTL 60

DB 1 MSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGDATYGKLTIKFKICTTGKLPVPWPTL 60

XX OS Unidentified.
 XX PN WO200240528-A1.
 XX PD 23-MAY-2002.
 XX PF 26-OCT-2001; 2001WO-CN01506.
 XX PR 26-OCT-2000; 2000CN-0125859.
 XX PA (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.
 XX PI Lu C, Huang J, Zhao Y, Zhang F, Chen X;
 XX DR WPI; 2002-427084/45.
 XX PS Method for producing non-natural silk by Bombyx mori with modifying
 PT technology and light chains of fibroin in natural silk by DNA recombination
 PT technology and protein engineering, applicable in sericulture and
 PT textile industry -
 XX Claim 4; Page 40-42; 50pp; Chinese.
 XX This invention relates to a silk which constitutes recombinant Bombyx
 CC mori silk fibroin obtained by modification or variation at the heavy
 CC or light chains of the fibroin. The method is for producing non-natural
 CC silk, which is applicable in sericulture and textile industry. Such
 CC modified silk has improved performance. The present sequence is a
 CC protein related to the invention.
 XX Sequence 655 AA;
 SQ
 Query Match 52.9%; Score 1353; DB 23; Length 655;
 Best Local Similarity 86.0%; Pred. No. 1.7e-105;
 Matches 264; Conservative 4; Mismatches 19; Indels 20; Gaps 5;
 QY 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVPWPTL 60
 DB 6 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVPWPTL 65
 QY 61 VTTLTYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 66 VTTLTYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 125
 QY 121 NRLELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
 DB 126 NRLELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 185
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYK-- 238
 DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYKGI 245
 QY 239 --SGSGSG--GGSGSGSG-----SGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 281
 DB 246 PADGRSGGLGGGPGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 304
 QY 282 GSGGLRS 288
 DB 305 GTGGLGS 311
 RESULT 8
 AAW85012
 ID AAW85012 standard; Protein; 719 AA.
 AC AAW85012;
 XX
 XX 08-FEB-1999 (first entry)
 XX Smad2-green fluorescent protein fusion product.
 DE Human; Smad2 gene; fusion protein; green fluorescent protein; GFP;
 KW

KW intracellular signalling; chimera.
 XX Chimeric - Aequorea victoria.
 OS Chimeric - Homo sapiens.
 XX WO9845704-A2.
 XX 15-OCT-1998.
 XX 07-APR-1998; 98WO-DK00145.
 XX 07-APR-1997; 97DK-0000392.
 XX (NOVO) NOVO-NORDISK AS.
 XX Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;
 PI Tullin S;
 XX WPI; 1998-594491/50.
 DR N-PSDB; AAV71027.
 XX Determining effect on signalling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents
 PS Example 5; Pages 102-103; 326pp; English.
 XX The present sequence represents a human Smad2-green fluorescent
 CC protein fusion product. The fusion protein is used in an assay
 CC that exemplifies the invention. The specification describes how
 CC quantitative information about the influence of a molecule on a cellular
 CC response is obtained by recording the variation, caused by the molecule,
 CC on mechanically intact living cells, in the spatially distributed light
 CC emitted from a luminophore present in the cells. The variation in light
 CC distribution is processed to provide information that correlates spatial
 CC identity agents that (in)directly affect intracellular signalling,
 CC especially to screen for potential therapeutic agents or toxins, and
 CC to identify new drug targets.
 XX Sequence 719 AA;
 SQ
 Query Match 50.6%; Score 1295; DB 19; Length 719;
 Best Local Similarity 81.0%; Pred. No. 1.5e-100;
 Matches 251; Conservative 9; Mismatches 36; Indels 14; Gaps 1;
 QY 1 MSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVPWPTL 60
 DB 2 VSKGEELFTGVVPILVELDGVNKGKFSVSGEGDATYKLTLPKICTTGKLPVPWPTL 61
 QY 61 VTTLTYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 62 VTTLTYGVOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
 QY 121 NRLELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
 DB 122 NRLELKGIDFKEDGNILGHKLEYNHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 181
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYKSG 240
 DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHVMVLLFEVTAAGITHGMDELYKSG 241
 QY 241 GSG 286
 DB 242 LRSRAQASNTWSSILPFTPPVVKLLGWKKSAGSGGAGGGEQKQEKVAKVSLV 301
 QY 287 RSVKLTSDFD 296
 DB 302 KKLKKTGRLD 311
 RESULT 9

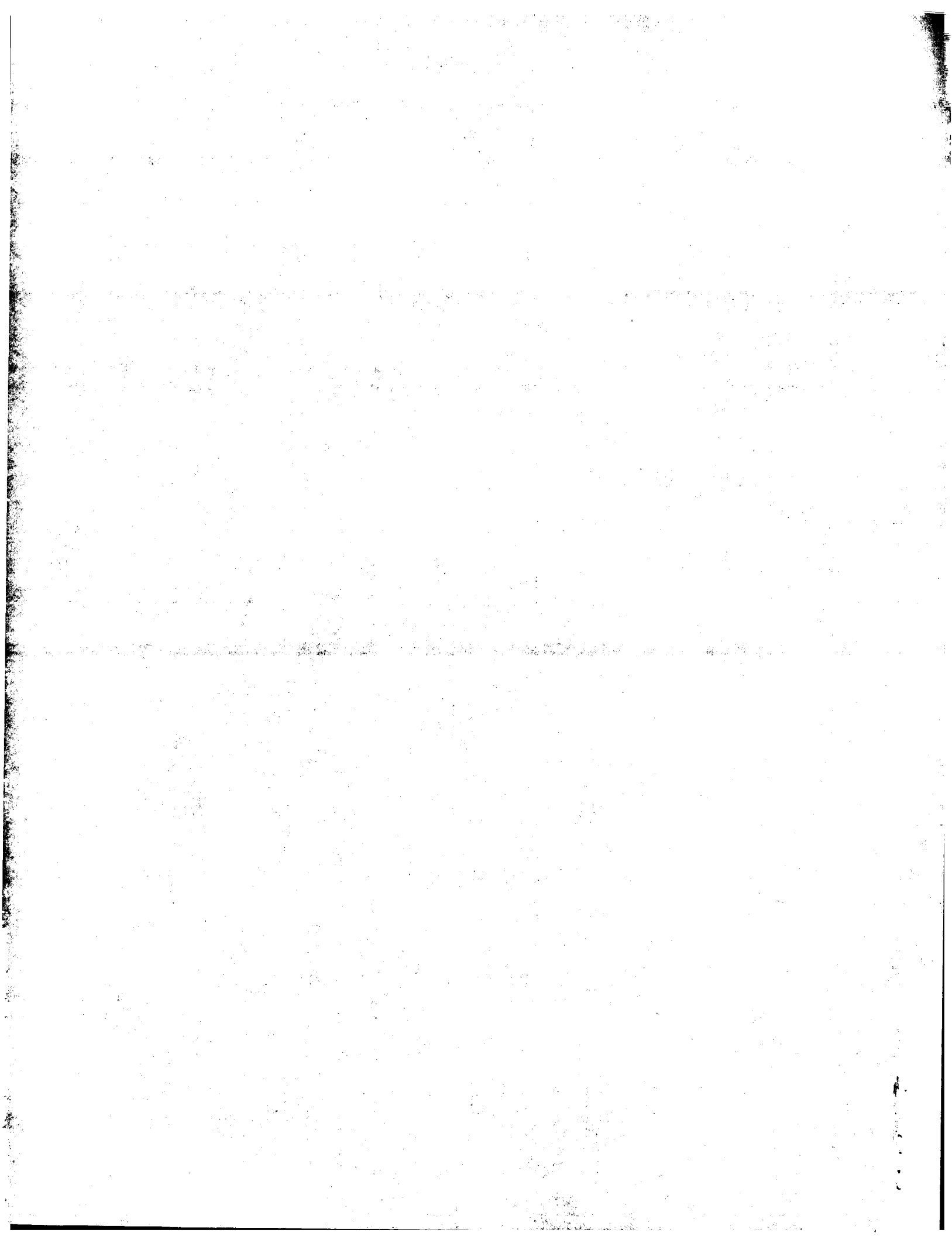
AAV70779

ID AAY70779 standard; Protein: 719 AA.
 AC AAY70779;
 XX
 DT 31-JUL-2000 (first entry)
 XX
 DE EGFP-SMAD2 fusion protein construct.
 XX
 KW Fusion construct; EGFP-SMAD2; Green fluorescent protein; cytokine;
 KW cellular response; luminophore; screening; medicament; signal transducer;
 KW light/fluorescence intensity; intracellular signalling pathway; mutant;
 KW fluorescent probe; signal transduction; transforming growth factor-beta;
 KW TGF-beta.
 XX
 OS Chimeric - Homo sapiens.
 OS Chimeric - Aequorea victoria.
 XX
 PN WO200023615-A2.
 XX
 PD 27-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-DK00562.
 XX
 PR 15-OCT-1999; 98DK-0001320.
 XX
 PA (BIOI-) BIOIMAGE AS.
 XX
 PI Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;
 PI Hagel G;
 XX
 DR WPI; 2000-339710/29.
 DR N-PSDB; AAD00037.
 XX
 XX Quantifying cellular response to an influence, useful for detecting
 PT intracellular translocation or redistribution of biologically active
 PT substances comprises recording changes in spatially distributed light
 PT emitted from a luminophore -
 XX
 PS Example 4; Page 110-112; 150pp; English.
 XX
 CC The patent discloses a method for extracting quantitative information
 CC relating to an influence on a cellular response in mechanically intact
 CC living cells, by recording variation in spatially distributed light
 CC emitted from a luminophore, as a change in light/fluorescence intensity.
 CC The luminophore present in the cells, is capable of being redistributed
 CC in a manner which is related with the degree of the influence. This
 CC method is useful as a screening program, for the identification of a
 CC biologically active substance, that directly or indirectly affects an
 CC intracellular signalling pathway. It is also potentially useful as a
 CC medicament. The fluorescent probe is useful in the back-tracking of
 CC signal transduction pathways. The present sequence is a fusion protein
 CC construct EGFP-SMAD2, comprising the human Smad2 gene, fused to a
 CC derivative of the luminous green fluorescent protein, EGFP. Smad2 is
 CC a signal transducer, that is induced by some members of the transforming
 CC growth factor-beta (TGF) family of cytokines. This construct is under the
 CC control of a CMV promoter and functions as the luminophore, that is
 CC useful to monitor signalling pathways and identify compounds that
 CC modulate the pathways in living cells.
 XX
 SQ Sequence 719 AA;
 Query Match 50.6%; Score 1295; DB 21; Length 719;
 Best Local Similarity 81.0%; Pred. No. 1.5e-100;
 Matches 251; Conservative 9; Mismatches 36; Indels 14; Gaps 1;
 QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKTLKFLICTTGKLPVPWPTL 60
 DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKTLKFLICTTGKLPVPWPTL 61
 QY 61 VTTLTYGVQCFSRYPDHMKQKQDFFKSAMPEGVYQVETIFFKDDGNYKTRAEVKFECDTLV 120
 DB 62 VTTLTYGVQCFSRYPDHMKQKQDFFKSAMPEGVYQVETIFFKDDGNYKTRAEVKFECDTLV 121

QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHINIEDSGVOLAD 180
 DB 122 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHINIEDSGVOLAD 181
 QY 181 HYQONTPTGDGPVLLPDNHYLSTQSALSADPNKRDHNVLLFEVTAAGTTHGMDELYKSG 240
 DB 182 HYQONTPTGDGPVLLPDNHYLSTQSALSADPNKRDHNVLLFEVTAAGTTHGMDELYKSG 241
 QY 241 GSGSGGSG 286
 DB 242 LRSRAQASNTMSSILPFTPPVVKLLGHWKKSAGSGGAGGGEQNGQEKWCEKAVKSLV 301
 QY 287 RSVKLTSDFD 296
 DB 302 KKLKKTGRLD 311
 RESULT 10
 ABB08620
 ID ABB08620 standard; Protein: 356 AA.
 XX
 AC ABB08620;
 XX
 DT 10-APR-2002 (first entry)
 XX
 DE Yeast polypeptide 1.
 XX
 KW Yeast; drug; cell wall; GPI anchor protein.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN WO200183733-A1.
 XX
 PD 08-NOV-2001.
 XX
 PF 26-APR-2001; 2001WO-JP03630.
 XX
 PR 01-MAY-2000; 2000JP-0132041.
 XX
 PA (DAUC) DAIICHI PHARM CO LTD.
 XX
 PI Kitamura A, Someya K, Nakajima R;
 DR WPI; 2002-097496/13.
 DR N-PSDB; ABA97833.
 XX
 PT Screening for drugs that act on cell walls, involves culturing
 PT microorganisms with a reporter protein acting as a GPI anchor protein
 PT in their cell walls -
 XX
 PS Disclosure; Fig 4-5; 44pp; Japanese.
 XX
 CC The invention relates to screening for drugs that act on cell walls,
 CC comprising culturing microorganisms with a reporter protein acting as a
 CC GPI anchor protein in their cell walls in the presence of the test
 CC substance and assaying the amount of reporter protein produced in the
 CC culture. The present sequence is that of a polypeptide useful to the
 CC invention.
 XX
 SQ Sequence 356 AA;
 Query Match 50.5%; Score 1293.5; DB 23; Length 356;
 Best Local Similarity 88.0%; Pred. No. 7.8e-101;
 Matches 250; Conservative 3; Mismatches 26; Indels 5; Gaps 2;
 QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKTLKFLICTTGKLPVPWPTL 60
 DB 40 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKTLKFLICTTGKLPVPWPTL 99
 QY 61 VTTLTYGVQCFSRYPDHMKQKQDFFKSAMPEGVYQVETIFFKDDGNYKTRAEVKFECDTLV 120
 DB 100 VTTLTYGVQCFSRYPDHMKQKQDFFKSAMPEGVYQVETIFFKDDGNYKTRAEVKFECDTLV 159

PN WO200050872-A2.
XX 31-AUG-2000.
XX 25-FEB-2000; 200WO-US04794.
XX 26-FEB-1999; 99US-0122152.
PR 08-MAR-1999; 99US-0123399.
PR 12-JUL-1999; 99US-0352171.
XX (CELL-) CELLOMICS INC.
XX
XX
PI Giuliano KA, Kapur R;
XX
XX WPI; 2000-594086/56.
DR N-PSDB; AAA93351.
XX
XX Automated cell-based characterization of toxin by contacting cells
PT containing luminescent reporter molecules with test substance and
PT analyzing optically -
XX
XX Example 11; Page 173-174; 336pp; English.
XX
CC The invention relates to systems, methods and reagents for cell-based
CC screening or detection of compounds which affect particular biological
CC functions. The methods of the invention utilise fluorescent biodevator
CC molecules which, when acted on by a compound of interest, cause an
CC alteration in the cellular distribution of at least the fluorescent
CC moiety. In one embodiment, the biosensors comprise heat shock proteins
CC (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent
CC protein (GFP), or derivatives thereof). Such biosensors are located in
CC the cytoplasm, but on stress activation translocate to the nucleus. In
CC another embodiment biodevator proteins can be used to detect protease
CC activity. Such protease biodevator fusion proteins comprise one or more
CC fluorescent proteins; a recognition signal which is cleaved by the
CC protease; and at least one cellular localisation signal. The latter two
CC components may be components of a single protein which is acted upon by
CC the protease, or may be from heterologous sources. Due to the
CC localisation signal, the biodevator protein is localised to a
CC particular region of the cell. Once acted on by the protease of interest,
CC the fluorescent protein is cleaved from the localisation sequence, and
CC is free to migrate to other locations within the cell. The presence of a
CC second localisation signal attached to the fluorescent protein enables
CC the fluorescent protein to be directed to a different cellular
CC compartment after cleavage of the protease recognition sequence. The
CC change in distribution of the fluorescent protein can be detected using
CC imaging methods with a high degree of spatial resolution. The methods
CC and biosensors of the invention can be used to investigate a wide range
CC of cellular activities and to screen compounds which modulate these
CC activities. Biosensors containing a recognition site for caspase, for
CC example, may be used for the screening of compounds which modulate
CC apoptosis, while biosensors containing other protease recognition sites
CC may be used for the detection of proteolytic toxins (such as anthrax
CC lethal factor). The method provides improved target validation and
CC candidate compound optimisation by combining many cell screening formats
CC with fluorescence-based molecular reagents and computer-based feature
CC extraction, data analysis and automation, resulting in increased
CC quantity and speed of data collection and faster evaluation of drug
CC candidates. Sequences AAB22860-B22876 and AAB22936-B22941 represent
CC biosensor fusion proteins produced in an exemplification of the
CC invention.
XX
XX Sequence 294 AA;
SQ
Query Match 50.3%; Score 1288.5; DB 21; Length 294;
Best Local Similarity 91.0%; Pred. No. 1.6e-100;
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;
QY 1 MSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGLTKLFCITGKLPVPWPTL 60
DB :|||||
2 VSKGEELFTGVVPIVLVDGVDNGHKFSVSGEGGDATYGLTKLFCITGKLPVPWPTL 61
QY 61 VTTLTYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 120

Db 62 VTTLTYGQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFGDTLV 121
QY 121 NRIELKGDIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKANFKIRHINIEDGSGVOLAD 180
Db 122 NRIELKGDIDFKEDGNILGHKLEYNNSHNYYIMADKQKNGIKANFKIRHINIEDGSGVOLAD 181
QY 181 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYKSG 240
Db 182 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMYLLEFVTAAGTTHGMDELYKSG 241
QY 241 GSISGGQSGGSGG-QSGGSGSGGSGG 267
Db 242 LRSGAGAGAGAGAGADEVDGAGADEVDG 269
Search completed: March 13, 2003, 15:03:39
Job time : 61.833 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:02:34 ; Search time 22.6281 Seconds
(without alignments)
620.235 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGEELFTGVVPILVELDG.....LGEWYTMDPACKLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1290.5	50.4	890	4	US-09-513-783A-174
2	1288.5	50.3	294	4	US-09-513-783A-2
3	1286	50.2	1070	4	US-09-091-042A-2
4	1283	50.1	459	4	US-09-513-783A-170
5	1270	49.6	238	3	US-08-893-327-16
6	1270	49.6	247	3	US-08-893-327-18
7	1269	49.6	783	4	US-09-513-783A-176
8	1267	49.5	941	4	US-09-513-783A-172
9	1266	49.5	1452	4	US-09-127-227-2
10	1265.5	49.4	350	4	US-09-513-783A-30
11	1265	49.4	805	4	US-09-513-783A-178
12	1265	49.4	1407	4	US-08-974-549A-628
13	1263	49.3	302	4	US-09-513-783A-18
14	1262	49.3	238	1	US-08-337-915A-2
15	1262	49.3	238	4	US-09-121-539-1
16	1262	49.3	238	5	PCT-US95-14692-2
17	1259.5	49.2	642	2	US-08-818-253-2
18	1259.5	49.2	642	4	US-08-818-252-2
19	1259.5	49.2	652	2	US-08-818-253-4
20	1259.5	49.2	652	4	US-08-818-252-4
21	1259	49.2	295	4	US-09-513-783A-20
22	1258	49.1	238	1	US-08-753-143-2
23	1258	49.1	238	2	US-08-679-865-2
24	1258	49.1	238	2	US-08-680-876-2
25	1258	49.1	238	2	US-08-792-553-2
26	1258	49.1	238	3	US-08-753-144-2
27	1258	49.1	238	4	US-09-094-359-2

28	1258	49.1	238	4	US-09-172-063-2	Sequence 2, Appl1
29	1258	49.1	238	4	US-09-263-975-2	Sequence 2, Appl1
30	1258	49.1	238	4	US-08-727-452-2	Sequence 2, Appl1
31	1258	49.1	238	4	US-09-418-785-1	Sequence 1, Appl1
32	1257	49.1	292	4	US-09-513-783A-16	Sequence 16, Appl1
33	1257	49.1	347	4	US-09-513-783A-28	Sequence 28, Appl1
34	1256.5	49.1	642	2	US-08-818-253-6	Sequence 6, Appl1
35	1256.5	49.1	642	4	US-08-818-252-6	Sequence 6, Appl1
36	1256.5	49.1	656	2	US-08-818-253-8	Sequence 8, Appl1
37	1256.5	49.1	656	4	US-08-818-252-8	Sequence 8, Appl1
38	1256	49.1	238	3	US-08-643-704A-49	Sequence 49, Appl1
39	1256	49.1	281	4	US-09-062-102-1	Sequence 1, Appl1
40	1256	49.1	281	4	US-09-364-946-1	Sequence 1, Appl1
41	1255	49.0	239	4	US-09-172-063-3	Sequence 3, Appl1
42	1255	49.0	239	4	US-09-513-783A-46	Sequence 46, Appl1
43	1255	49.0	239	4	US-09-316-919-4	Sequence 4, Appl1
44	1255	49.0	282	4	US-09-513-783A-14	Sequence 14, Appl1
45	1255	49.0	296	4	US-09-513-783A-12	Sequence 12, Appl1

ALIGNMENTS

RESULT 1
US-09-513-783A-174
; Sequence 174, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513/783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 174
; LENGTH: 890
; TYPE: PNT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSC70
US-09-513-783A-174

Query Match		50.4%	Score 1290.5;	DB 4;	Length 890;
Best Local Similarity		64.4%	Pred No. 5e-100;		
Matches 282;		Conservative 22;	Mismatches 77;	Indels 57;	Gaps 12;
QY	1	MSKGEELFTGVVPILVELDGVN	GKHFVS	SGEGDATY	GKLT
DB	2	VSKGEELFTGVVPILVELDGVN	GKHFVS	SGEGDATY	GKLT
QY	61	VTTLYGVQCSRYPDHMKQHDFF	KSAMPEGV	YQERTIFF	KDDGNYKTRAEV
DB	62	VTTLYGVQCSRYPDHMKQHDFF	KSAMPEGV	YQERTIFF	KDDGNYKTRAEV
QY	121	NRIELKGI	DFKDDGNILGHKLEYN	NSHNYIMADKQNGIKAN	FEKIRHINIEDG
DB	122	NRIELKGI	DFKDDGNILGHKLEYN	NSHNYIMADKQNGIKAN	FEKIRHINIEDG
QY	181	HYQONTPI	GDGPVLLPDNNHYLSTQ	SALSKDPNEKRDHVMVLL	FEVTAAGITHOMDELYKSG
DB	182	HYQONTPI	GDGPVLLPDNNHYLSTQ	SALSKDPNEKRDHVMVLL	FEVTAAGITHOMDELYKSG
QY	241	-GSGGSGGSGGSGGSGGSGG	SGGSGGSGGSGGSGGSGGSGG	SGGSGGSGGSGGSGGSGGSGG	SGGSGGSGGSGGSGGSGGSGG
DB	242	LKSMKGP	AVGIDLTGTTYS	CVGV-FQHGKVEI	ANDQGNRTTPSYVAF
QY	284	SGLSRVKLT	SDPDNPRWTGRH	-----	KHMFENLFDVNHNGKISL
DB	301	NQVAMPNT	VFDAKRLIGRRFDD	AVOSDMKH-WPFV	VVNDAGRPKV-QVEYK
QY	334	NNLGAT	----	PEQAK-----	RHKDAVEAF

```
Db 353 ---GETKSYPEEVSWMVTKKEIAEALGKTVNAVTVV--PAYFNDSQRAKRDAGT 407
QY 385 YAK-----NEPTLIRI 395
Db 408 IAGLNVLRIINEPTAAAI 425

RESULT 2
US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match 50.3%; Score 1288.5; DB 4; Length 294;
Best Local Similarity 91.0%; Pred. No. 1.6e-100;
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYGLTKLKFICTTGKLPVPWPTL 60
Db 2 VSKGEELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYGLTKLKFICTTGKLPVPWPTL 61
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 122 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITTHGMDELYKSG 240
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITTHGMDELYKSG 241
QY 241 GSGSGGSGSGSGG-QSGSGSGSGSGG 267
Db 242 LRSAGAGAGAGAGAGADEVDGAGADEVDG 269

RESULT 3
US-09-091-042A-2
; Sequence 2, Application US/09091042A
; Patent No. 6455300
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America
; as represented by the Secretary
; Department of Health and Human Services
; Washington, D.C.
; Hager Ph.D., Gordon L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MONITORING
; DNA BINDING MOLECULES IN LIVING CELLS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
```

```
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,042A
; FILING DATE: 08-Jun-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/008,373
; FILING DATE: 08 Dec 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Selby, Elizabeth
; REGISTRATION NUMBER: 38298
; REFERENCE/DOCKET NUMBER: 14014.0183
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-091-042A-2

Query Match 50.2%; Score 1286; DB 4; Length 1070;
Best Local Similarity 97.6%; Pred. No. 1.6e-99;
Matches 239; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYGLTKLKFICTTGKLPVPWPTL 60
Db 23 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGEGGDATYGLTKLKFICTTGKLPVPWPTL 82
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 83 VTTLYGVQCFSRYPDHMKQHDFFKFSAMPEGVYQVQERTIFFKDDGNYKTRAEVKFEGDTLV 142
QY 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
Db 143 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQNGIKVNFKIRHNIEDGSVQLAD 202
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITTHGMDELYKSG 240
Db 203 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLVLEFVTAAGITTHGMDELYKSG 262
QY 241 GSGSG 245
Db 263 GAGAG 267

RESULT 4
US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```

: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-09-513-783A-170

Query Match          50.1%; Score 1283; DB 4; Length 459;
Best Local Similarity 66.8%; Pred. No. 8.5e-100;
Matches 266; Conservative 11; Mismatches 55; Indels 66; Gaps

QY 1 MSGGEELFTGVPLILVELDGVNKHFSVSGEGEGDATYKGLTLKEICTITGKLPVWPPTL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 VSGEELFTGVPLILVELDGVNKHFSVSGEGEGDATYKGLTLKEICTITGKLPVWPPTL 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNKTKRAEVKFECDTLV 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 NRTELKGIDPKEDGNILGHKLEYNINSHNVYIMADKQKNGIKRANFIRINIEDGVSQVLAD 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 NRTELKGIDPKEDGNILGHKLEYNINSHNVYIMADKQKNGIKRANFIRINIEDGVSQVLAD 181
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 HYQONTPIGDPVLLPDNHVLSYTSQALS KDPNKRDHVMVLEEVTAAGITHGMDLYKSG 240
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 HYQONTPIGDPVLLPDNHVLSYTSQALS KDPNKRDHVMVLEEVTAAGITHGMDLYKSG 241
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 GSSGGSGGSGSGS-----GGQS-----G-----G 258
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 242 LRSRAASRAESASMTERRVPFSLLRGSPWDPPFRDWTYPSHSLFDQAFGLPRLPEWSQWLQ 301
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 259 GSGSGG-----QSGSGSGSGSGSGSGSGSLRSVKLTSPDNPWTRIGRHKHF 308
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 GSSWPGVYRPLPAAATESPAVAAPYSRALSRQLSSGVSEIRITAD----RWRVS----- 352
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 309 NFDLVNHNGKISLDEMYYKASDIVINNLCATPQAKRH 346
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 353 --LDVNH---FAPDELVTKDKGWETTKGHEERODEH 385
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

: TELEFAX: (713) 789-2679
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-893-327-16

Query Match 49.6%; Score 1270; DB 3; Length 238;
Best Local Similarity 99.6%; Pred. No. 4.2e-99;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVELDGVNGHKFSVSGEGSDATYGKLTLLKFTCTTGKLPVPWPTL 60
   |||||
Db 1 MSKGEELFTGVVPLVELDGVNGHKFSVSGEGSDATYGKLTLLKFTCTTGKLPVPWPTL 60

QY 61 VTTITYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTLV 120
   |||||
Db 61 VTTITYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNGYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHHYIMADKOKNGIKANFKIRHNIEDGSVOLAD 180
   |||||
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHHYIMADKOKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQQTPTIGDGPVLLPDNHYLSQTSAISKDPNEKRDHMVLLFEVTAAGITHGMDELYK 238
   |||||
Db 181 HYQQTPTIGDGPVLLPDNHYLSQTSAISKDPNEKRDHMVLLFEVTAAGITHGMDELYK 238

RESULT 6
US-08-893-327-18
: Sequence 18, Application US/08893327
: Patent No. 6020192
: GENERAL INFORMATION:
: APPLICANT: Zolotukhin, Sergei
: APPLICANT: Hauswirth, William W.
: APPLICANT: Muzyczka, Nicholas
: TITLE OF INVENTION: Humanized Green Fluorescent Protein
: TITLE OF INVENTION: Genes and Methods
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/893,327
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/588,201
: FILING DATE: 18-JAN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Kitchell, Barbara S.
: REGISTRATION NUMBER: 33,928
: REFERENCE/DOCKET NUMBER: UFLA:062\KIT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000
: TELEFAX: (713) 789-2679
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 247 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein

```

US-08-893-327-18

Query Match 49.6%; Score 1270; DB 3; Length 247;
Best Local Similarity 99.6%; Pred. No. 4.4e-99;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
DB 10 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 69
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 70 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 129
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKAKFKIRHNIEDGSVQLAD 180
DB 130 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 189
QY 181 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYK 238
DB 190 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYK 247

RESULT 7

US-09-513-783A-176
; Sequence 176, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 176
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSF1
US-09-513-783A-176

Query Match 49.6%; Score 1269; DB 4; Length 783;
Best Local Similarity 91.0%; Pred. No. 2.7e-98;
Matches 242; Conservative 3; Mismatches 21; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 61
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 62 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKAKFKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYKSG 241
QY 241 GSGGG 266
DB 242 LRSRAQASNSAVEMDLPVGPAGAGPS 267

RESULT 8

US-09-513-783A-172
; Sequence 172, Application US/09513783A
; Patent No. 6416959

; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 172
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP70
US-09-513-783A-172

Query Match 49.5%; Score 1267; DB 4; Length 941;
Best Local Similarity 97.6%; Pred. No. 5.1e-98;
Matches 239; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 61
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 62 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKAKFKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKVNFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTQALSADPNKRDHVVLEFVTAAGITHGMDELYKSG 241
QY 241 GSGSG 245
DB 242 MSVVG 246

RESULT 9

US-09-127-227-2
; Sequence 2, Application US/09127227
; Patent No. 6399354
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/09/127,227
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-09-127-227-2

Query Match 49.5%; Score 1266; DB 4; Length 1452;
Best Local Similarity 99.2%; Pred. No. 1.1e-97;
Matches 236; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 60
DB 1215 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKGLTLKFICTTGKLPVPWPTL 1274
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAPEGVYQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120

```
Db 1275 VTTFTYGVOCFSRYPDHMKQHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 1334
Qy 121 NRIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLAD 180
Db 1335 NRIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLAD 1394
Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 238
Db 1395 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYK 1452

RESULT 10
US-09-513-783A-30
; Sequence 30, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: NLS-Fred25-cellubrevin construct
US-09-513-783A-30

Query Match 49.4%; Score 1265.5; DB 4; Length 350;
Best Local Similarity 93.4%; Pred. No. 1.7e-98;
Matches 239; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy 2 SKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKLTCLKFICTTGKLPVPMPTLV 61
Db 9 SKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKLTCLKFICTTGKLPVPMPTLV 68
Qy 62 TLTLYGVOCFSRYPDHMKQHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Db 69 TLTLYGVOCFSRYPDHMKQHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 128
Qy 122 RIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLADH 181
Db 129 RIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLADH 188
Qy 182 YQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYKSGG 241
Db 189 YQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYNTCM 248
Qy 242 SGGSGGSGSGSGGQS 257
Db 249 S-TGVPSSGSAATGSN 263

RESULT 11
US-09-513-783A-178
; Sequence 178, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 178
; LENGTH: 805
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-NFKB
US-09-513-783A-178

Query Match 49.4%; Score 1265; DB 4; Length 805;
Best Local Similarity 98.8%; Pred. No. 6.1e-98;
Matches 237; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKLTCLKFICTTGKLPVPMPTLV 60
Db 2 VSKGEELFTGVVPILVELDGVNKHFSVSGEGDATYKLTCLKFICTTGKLPVPMPTLV 61
Qy 61 VTTLYGVOCFSRYPDHMKQHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 62 VTTLYGVOCFSRYPDHMKQHDFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Qy 121 NRIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLKYKNIKANFIRHNIEDGSGVOLAD 181
Qy 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYKSG 240
Db 182 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLEFVTAAGITHGMDELYKSG 241

RESULT 12
US-08-974-549A-628
; Sequence 628, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
```


Query Match 49.3%; Score 1262; DB 1; Length 238;
Best Local Similarity 98.7%; Pred. No. 2e-98;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYGLTLKFLICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYGLTLKFLICTTGKLPVPWPTL 60
QY 61 VTTLTGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTFSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 15
US-09-121-539-1
; Sequence 1, Application US/09121539B
; Patent No. 6194548
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshio
; APPLICANT: Tsukamoto, No. 61945481yo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; FILE REFERENCE: 046124-5005
; CURRENT APPLICATION NUMBER: US/09/121,539B
; CURRENT FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: JP 026418/1998
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; OTHER INFORMATION: Green fluorescent protein
US-09-121-539-1

Query Match 49.3%; Score 1262; DB 4; Length 238;
Best Local Similarity 98.7%; Pred. No. 2e-98;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYGLTLKFLICTTGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFVSQGEQDATYGLTLKFLICTTGKLPVPWPTL 60
QY 61 VTTLTGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTFSYGVQCFSPYDPMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITHGMDELYK 238

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:05:14 ; Search time 19.9127 Seconds
(without alignments)
1104.115 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2360
Sequence: 1 MSKGEELFTGVVPILVELDG.....LGFWYTMDPACKLYGGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues
Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	49.5	1452	12	US-10-050-673-2
2	1262	49.3	238	9	US-09-866-538-2
3	1262	49.3	238	9	US-09-900-345A-125
4	1262	49.3	238	10	US-09-920-922-4
5	1262	49.3	238	10	US-09-852-000-1
6	1259.5	49.2	642	9	US-09-554-000-2
7	1259.5	49.2	652	9	US-09-554-000-4
8	1259	49.2	243	9	US-09-900-345A-60
9	1259	49.2	243	9	US-09-900-345A-62
10	1259	49.2	243	9	US-09-900-345A-64
11	1259	49.2	243	9	US-09-900-345A-66
12	1259	49.2	243	9	US-09-900-345A-68
13	1259	49.2	243	9	US-09-900-345A-70
14	1258	49.1	238	9	US-10-057-505-2
15	1258	49.1	238	10	US-09-884-681-2
16	1258	49.1	238	12	US-10-024-686-2
17	1258	49.1	243	9	US-09-900-345A-54
18	1258	49.1	243	9	US-09-900-345A-56
19	1258	49.1	243	9	US-09-900-345A-58

20	1258	49.1	243	9	US-09-900-345A-114	Sequence 114, App
21	1258	49.1	243	9	US-09-900-345A-116	Sequence 116, App
22	1258	49.1	243	9	US-09-900-345A-118	Sequence 118, App
23	1258	49.1	243	9	US-09-900-345A-120	Sequence 120, App
24	1257	49.1	243	9	US-09-900-345A-2	Sequence 2, Appli
25	1257	49.1	243	9	US-09-900-345A-4	Sequence 4, Appli
26	1257	49.1	243	9	US-09-900-345A-6	Sequence 6, Appli
27	1257	49.1	243	9	US-09-900-345A-8	Sequence 8, Appli
28	1257	49.1	243	9	US-09-900-345A-10	Sequence 10, Appl
29	1257	49.1	243	9	US-09-900-345A-12	Sequence 12, Appl
30	1257	49.1	243	9	US-09-900-345A-14	Sequence 14, Appl
31	1257	49.1	243	9	US-09-900-345A-16	Sequence 16, Appl
32	1257	49.1	243	9	US-09-900-345A-18	Sequence 18, Appl
33	1257	49.1	243	9	US-09-900-345A-20	Sequence 20, Appl
34	1257	49.1	243	9	US-09-900-345A-22	Sequence 22, Appl
35	1257	49.1	243	9	US-09-900-345A-24	Sequence 24, Appl
36	1257	49.1	243	9	US-09-900-345A-26	Sequence 26, Appl
37	1257	49.1	243	9	US-09-900-345A-28	Sequence 28, Appl
38	1257	49.1	243	9	US-09-900-345A-30	Sequence 30, Appl
39	1257	49.1	243	9	US-09-900-345A-32	Sequence 32, Appl
40	1257	49.1	243	9	US-09-900-345A-34	Sequence 34, Appl
41	1257	49.1	243	9	US-09-900-345A-36	Sequence 36, Appl
42	1257	49.1	243	9	US-09-900-345A-38	Sequence 38, Appl
43	1257	49.1	243	9	US-09-900-345A-40	Sequence 40, Appl
44	1257	49.1	243	9	US-09-900-345A-42	Sequence 42, Appl
45	1257	49.1	243	9	US-09-900-345A-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-10-050-673-2
; Sequence 2, Application US/10050673
; Patent No. US20020151033A1
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; FILE REFERENCE: H098-05
; CURRENT APPLICATION NUMBER: US/10/050.673
; CURRENT FILING DATE: 2002-01-16
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PFT
; ORGANISM: herpesvirus
US-10-050-673-2

Query Match	49.5%	Score 1266;	DB 12;	Length 1452;
Best Local Similarity	99.2%	Pred. No. 3e+80;		
Matches	236;	Conservative	0;	Mismatches 2;
Indels	0;	Gaps	0;	
Qy	1	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYTKLTKFICTTGKLPVPWPTL	60	
Db	1215	MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDGYTKLTKFICTTGKLPVPWPTL	1274	
Qy	61	VTTLTGYQCFSRYPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV	120	
Db	1275	VTTFTYGVQCFSRYPDHMKQHDFFKSAPEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV	1334	
Qy	121	NRIELKGIDFEDKNILGHKLEYNNSHVYIMADKQNGKANKFKIRHNIEDGSVQLAD	180	
Db	1335	NRIELKGIDFEDKNILGHKLEYNNSHVYIMADKQNGKANKFKIRHNIEDGSVQLAD	1394	
Qy	181	HYQQNTPIDGDPVLLPDNNHLYSTQSAISKDPNEKRDHVVLEFVTAAGITHGMDELYK	238	
Db	1395	HYQQNTPIDGDPVLLPDNNHLYSTQSAISKDPNEKRDHVVLEFVTAAGITHGMDELYK	1452	

```
RESULT 2
US-09-866-538-2
; Sequence 2, Application US/09866538
; Publication No. US20030032089A1
; GENERAL INFORMATION:
; APPLICANT: TSTEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-866-538-2

Query Match          49.3%; Score 1262; DB 9; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
;
Qy 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
;
Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
;
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
;
RESULT 3
US-09-900-345A-125
; Sequence 125, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-50S
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP humanized
; OTHER INFORMATION: control
; US-09-900-345A-125

Query Match          49.3%; Score 1262; DB 9; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
;
Qy 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
;
Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
;
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
;
RESULT 4
US-09-920-922-4
; Sequence 4, Application US/09920922
; Patent No. US20020083488A1
; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-920-922-4

Query Match          49.3%; Score 1262; DB 10; Length 238;
Best Local Similarity 98.7%; Pred. No. 6.9e-81;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVVPIVLVDGVDVNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTL 60
;
Qy 61 VTTLTYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
Db 61 VTTFSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
;
Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
;
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
;
RESULT 5
US-09-852-000-1
; Sequence 1, Application US/09852000
; Patent No. US20020099170A1
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshio
; APPLICANT: Tsukamoto, No. US20020099170A1yo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 046124-5005-US
; CURRENT APPLICATION NUMBER: US/09/852,000
```

[illegible]

```
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTC)5GFP
US-09-900-345A-62

Query Match          49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.le-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY   1 MSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 60
      : LSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 65
Db    6 LSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 65

QY   61 VTTLTYGVQCFSRYPDHMKQHDFKSAPEGYVQERTIFFKDGNKYTRAEVKFEGDTLV 120
      ||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    66 VTTYSYGVQCFSRYPDHMKQHDFKSAPEGYVQERTIFFKDGNKYTRAEVKFEGDTLV 125

QY   121 NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 180
      : NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 185
Db    126 NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 185

QY   181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDNHVLVEFYTAAGITHGMDLYK 238
      : HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDNHVLVEFYTAAGITHGMDLYK 243
Db    186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDNHVLVEFYTAAGITHGMDLYK 243

RESULT 10
US-09-900-345A-64
; Sequence 64, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; TITLE OF INVENTION: EFFICIENCY OF A CODON
; FILE REFERENCE: 10338-5U5
; CURRENT APPLICATION NUMBER: US/09/900,345A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU PP8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)5GFP
US-09-900-345A-64

Query Match          49.2%; Score 1259; DB 9; Length 243;
Best Local Similarity 98.3%; Pred. No. 1.le-80;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY   1 MSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 60
      : LSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 65
Db    6 LSKGEELFTGVVPILVELDGDVNGHKFSVSSEGEDATYGKLTLKFKICTTGKLPPVWPTL 65

QY   61 VTTLTYGVQCFSRYPDHMKQHDFKSAPEGYVQERTIFFKDGNKYTRAEVKFEGDTLV 120
      ||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    66 VTTYSYGVQCFSRYPDHMKQHDFKSAPEGYVQERTIFFKDGNKYTRAEVKFEGDTLV 125

QY   121 NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 180
      : NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 185
Db    126 NRIELKGIDFKEDGNLGHKLKLEYNINSHNVYIMADKQNKGIKANFKIRHNIEDGSVOLAD 185
```


RESULT 14

US-10-057-505-2
 ; Sequence 2, Application US/10057505
 ; Patent No. US20020164674A1
 ; GENERAL INFORMATION:
 ; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
 ; APPLICANT: AURORA BIOSCIENCES CORPORATION
 ; APPLICANT: TSIEN, Roger
 ; APPLICANT: HEIM, Roger
 ; APPLICANT: CUBITT, Andrew
 ; TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
 ; FILE REFERENCE: REG1260-3
 ; CURRENT APPLICATION NUMBER: US/10/057,505
 ; CURRENT FILING DATE: 2002-01-25
 ; PRIOR APPLICATION NUMBER: US 08/792,553
 ; PRIOR FILING DATE: 1997-01-31
 ; PRIOR APPLICATION NUMBER: US 09/396,003
 ; PRIOR FILING DATE: 1999-09-13
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Aequorea victoria
 ; US-10-057-505-2

Query Match 49.1%; Score 1258; DB 9; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1.3e-80;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFTCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFTCTTGKLPVWPPTL 60
 QY 61 VTLLTYGQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTSYGVQCFSRYPDHMKRHDFKFSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 15

US-09-884-681-2
 ; Sequence 2, Application US/09884681
 ; Patent No. US20020061546A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsiien, Roger Y.
 ; APPLICANT: Cubitt, Andrew B.
 ; TITLE OF INVENTION: Assays for Protein Kinases Using
 ; Fluorescent Protein Substrates
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/884,681
 ; FILING DATE: 19-Jun-2001
 ; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/679,865
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Storella, John S.
 ; REGISTRATION NUMBER: 32,944
 ; REFERENCE/DOCKET NUMBER: 023072-069000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 238 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 ; US-09-884-681-2

Query Match 49.1%; Score 1258; DB 10; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1.3e-80;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFTCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKLTILKFTCTTGKLPVWPPTL 60
 QY 61 VTLLTYGQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTSYGVQCFSRYPDHMKRHDFKFSAMPEGYVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKANFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHGMDELYK 238

Search completed: March 13, 2003, 15:12:35
 Job time : 23.9127 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:01:59 ; Search time 27.1537 Seconds
(without alignments)
1688.761 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2560
Sequence: 1 MSKGELFTGVVPILVELDG.....LGFWTMDPACELKYGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1256	49.1	238	1 JQ1514	green-fluorescent
2	1032	40.3	196	1 AQJENV	aequorin precursor
3	962	37.6	196	2 A26623	aequorin-l precursor
4	742	29.0	198	2 S39022	mitrocomin precurs
5	663	25.9	198	2 S28860	clytin - hydromedu
6	170	6.6	302	2 C84470	hypothetical prote
7	168	6.6	592	2 E82759	endo-1,4-beta-gluc
8	161	6.3	115	2 D61615	sericin MG-1 - gre
9	159.5	6.2	207	2 T07381	glycine-rich prote
10	158.5	6.2	268	1 C1HUL	calpain (EC 3.4.22
11	155.5	6.1	1226	2 T24045	hypothetical prote
12	155	6.1	641	1 Q0BE31	nuclear antigen EB
13	155	6.1	1218	2 E84537	hypothetical prote
14	154.5	6.0	622	2 I37984	keratin 9, type I,
15	154	6.0	1585	2 T31611	hypothetical prote
16	151.5	5.9	481	2 A35628	loricrin - mouse
17	149.5	5.8	204	2 T09592	protein corA, cold
18	149.5	5.8	221	2 T04592	glycine-rich cell
19	148.5	5.8	171	2 H84709	probable glycine-r
20	148	5.8	291	1 S31415	glycine-rich prote
21	148	5.8	1275	2 T49362	hypothetical prote
22	148	5.8	1901	2 T70806	hypothetical glyci
23	147.5	5.8	165	1 KNRZG1	glycine-rich cell
24	147.5	5.8	434	1 K23PIK	coat protein A - p
25	147	5.7	183	2 PNO109	keratin-like prote
26	147	5.7	526	1 K8ROVI	keratin, 54K type
27	147	5.7	569	1 KRMSEI	keratin, 59K type
28	147	5.7	995	2 T22942	hypothetical prote
29	146.5	5.7	266	1 CIRBL	calpain (EC 3.4.22

30	146	5.7	465	1 S01820	glycine-rich cell
31	145.5	5.7	166	1 KRBO2B	keratin, 68K type
32	145	5.7	263	2 A34466	calpain (EC 3.4.22
33	145	5.7	320	2 T09555	fibrillarlin - Arab
34	145	5.7	988	2 T08102	myrosinase-binding
35	144.5	5.6	266	1 C1PGL	calpain (EC 3.4.22
36	144.5	5.6	462	4 S33798	FUS/CHOP mutant fu
37	144	5.6	183	1 KNRZG2	glycine-rich cell
38	143.5	5.6	208	2 T46896	merozoite surface
39	143.5	5.6	434	2 S08091	gene III protein -
40	143.5	5.6	593	1 KRHU0	keratin 10, type I
41	143	5.6	150	2 C86224	hypothetical prote
42	143	5.6	420	2 A49642	transcription fact
43	142.5	5.6	167	2 S21359	keratin, type I, c
44	142.5	5.6	271	2 S34666	glycine-rich prote
45	142.5	5.6	528	2 G02127	fus-like protein -

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 23-Mar-2001
C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JQ1514; MUID:92175527; PMID:1347277
A:Accession: JS0692
A:Molecule type: DNA
A:Residues: 1-107 'S', 109-238 <PRAL>
A:Cross-references: GB:M62654; NID:g155662; PIDN:AAA27722.1; PID:g155663
A:Accession: JQ1514
A:Molecule type: mRNA
A:Residues: 1-99 'F', 101-140 'L', 142-218 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64,74-122,132-151;154-183;185-200 <PRA3>
R:Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; MUID:94364470; PMID:8082767
A:Accession: S48693
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-24 'Q', 26-156 'P', 158-171 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R:Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13 'V', 15-24 'Q', 26-44 'N', 46-153 'G', 155-156 'P', 158-171 'K', 173-227,
A:Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009
A:Experimental source: Clone gfp1
A:Accession: S51331
A:Molecule type: mRNA
A:Residues: 1-24 'Q', 26-29 'R', 31-83 'L', 85-153 'G', 155-156 'P', 158-171 'K', 173-208,
A:Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011
A:Experimental source: Clone gfp2
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65692; PDB:1GFL
A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 8
A>Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli
Nat. Biotechnol. 14, 1246-1251, 1996
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation; X-ray crystallography, 1.9 angstroms
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFN) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
 A:Gene: GFP
 A:Introns: 69/3; 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/Cross-link: 5-imidazolinone (Ser-Gly) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 49.1%; Score 1256; DB 1; Length 238;
 Best Local Similarity 97.5%; Pred. No. 3.7e-73;
 Matches 232; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGDYNGHKFSVSGEGEDATYKLTILKFICTTCKLPVPPPTL 60
 DB 1 MSKGEELFTGVVPIVLVELDGDYNGHKFSVSGEGEDATYKLTILKFICTTCKLPVPPPTL 60
 QY 61 VTTLTGVGOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFTGVGOCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRIELKIDFKEDGNILGHKLEYNSHNVIYIMADKOKNGTKANFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKIDFKEDGNILGHKMEYNVSHNVIYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
 QY 181 HYQONTPTGDGVPVLPDNNHLYSTQSALSKDPNEKRDRHMLLEFVTAAGITGHGMDLYK 238
 DB 181 HYQONTPTGDGVPVLPDNNHLYSTQSALSKDPNEKRDRHMLLEFVTAAGITGHGMDLYK 238

RESULT 2

AQJFN

aequorin precursor - hydromedusa (Aequorea victoria)
 C:Species: Aequorea victoria
 C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 26-May-2000
 C:Accession: A03020
 R:Inouye, S.; Nouguchi, M.; Sakaki, Y.; Takagi, Y.; Miyata, T.; Iwanaga, S.; Miyata, T.;
 Proc. Natl. Acad. Sci. U.S.A. 82, 3154-3158, 1985
 A:Title: Cloning and sequence analysis of cDNA for the luminescent protein aequorin.
 A:Reference number: A03020; MUID:85216460; PMID:3858813

A:Accession: A03020
 A:Molecule type: mRNA
 A:Residues: 1-196 <INO>
 A:Cross-references: GB:M11394; NID:g155658; PIDN:AAA27719.1; PID:g155659; GB:L29571; NID
 A:Experimental source: clone AQ440
 C:Comment: The precise function of residues 1-7 is not known.
 C:Comment: The authors suggest that there are three calcium-binding sites, and that res
 ts a blue light.
 C:Comment: Cysteine residues appear to be critical to the ability of this photoprotein t
 n, and mercaptoethanol is necessary to regenerate aequorin from apoaequorin.

C:Superfamily: calmodulin; calmodulin repeat homology
 F:8-196/Product: aequorin #status predicted <MAY>
 F:18-50/Domain: calmodulin repeat homology <EAF1>
 F:147-179/Domain: calmodulin repeat homology <EF2>
 F:31,33,35,37,42/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted
 F:124,126,128,130,135/Binding site: calcium (Asp, Asp, Asn, Ala, Glu) #status predicted
 F:160,162,164,166,171/Binding site: calcium (Asp, Asp, Ser, Glu, Glu) #status predicted

Query Match 40.3%; Score 1032; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 5.7e-59;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWGRHKKHMFNLDVNHNGKISLDEMVKASDIVINNLGATPEQAKRHK 347
 DB 7 SVKLTSDFDNPRWGRHKKHMFNLDVNHNGKISLDEMVKASDIVINNLGATPEQAKRHK 66
 QY 348 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407

Db 67 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 126
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
 Db 127 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 186
 QY 468 CEKLYGGAVP 477
 Db 187 CEKLYGGAVP 196

RESULT 3

A26623

aequorin-l precursor - hydromedusa (Aequorea victoria)
 C:Species: Aequorea victoria
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-May-2000
 C:Accession: A26623
 R:Prasher, D.C.; McCann, R.O.; Longiaru, M.; Cormier, M.J.
 Biochemistry 26, 1326-1332, 1987
 A:Title: Sequence comparisons of complementary DNAs encoding aequorin isoforms.
 A:Reference number: A26623; MUID:87185437; PMID:2882777

A:Accession: A26623
 A:Molecule type: mRNA
 A:Residues: 1-196 <PRA>
 A:Cross-references: GB:M16103; NID:g155652; PIDN:AAA27716.1; PID:g155653
 A:Note: the authors translated the codon GAT for residue 143 as Ala
 C:Comment: The precise function of residue 1-7 is not known.
 C:Superfamily: calmodulin; calmodulin repeat homology
 C:Keywords: calcium binding; EF hand; luminescence
 F:18-50/Domain: calmodulin repeat homology <EF1>
 F:111-143/Domain: calmodulin repeat homology <EF2>
 F:147-179/Domain: calmodulin repeat homology <EF3>

Query Match 37.6%; Score 962; DB 2; Length 196;
 Best Local Similarity 90.5%; Pred. No. 1.7e-54;
 Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWGRHKKHMFNLDVNHNGKISLDEMVKASDIVINNLGATPEQAKRHK 347
 Db 7 SVKLTSDFDNPRWGRHKKHMFNLDVNHNGKISLDEMVKASDIVINNLGATPEQAKRHK 66
 QY 348 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
 Db 67 DAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 126
 QY 408 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 467
 Db 127 QNGAITLDEWKATYKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMDDPA 186
 QY 468 CEKLYGGAVP 477
 Db 187 CEKLYGGAVP 196

RESULT 4

S39022

Mitrocomin precursor - Mitrocoma cellularia

C:Species: Mitrocoma cellularia

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-May-2000

C:Accession: S39022

R:Fagan, T.F.; Ohmura, Y.; Blinks, J.R.; Inouye, S.; Tsuji, F.I.

FEBS Lett. 333, 301-305, 1993

A:Title: Cloning, expression and sequence analysis of cDNA for the Ca(2+)-binding pho

A:Reference number: S39022; MUID:94039781; PMID:8224198

A:Accession: S39022

A:Molecule type: mRNA

A:Residues: 1-198 <FAG>

A:Cross-references: EMBL:L31623; NID:9468907; PIDN:AAA29298.1; PID:9468908

C:Superfamily: calmodulin; calmodulin repeat homology

C:Keywords: calcium binding; EF hand; luminescence

F:19-51/Domain: calmodulin repeat homology <EF1>

F:112-144/Domain: calmodulin repeat homology <EF2>

F:148-180/Domain: calmodulin repeat homology <EF3>

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C

R;Zurovec, M.; Sehnal, F.; Scheller, K.; Kumaran, A.K.

Insect Biochem. Mol. Biol. 22, 1992
A: Title: Silk gland specific cDNAs from *Galleria mellonella* L.
A: Reference number: A61615
A: Accession: D61615
A: Status: preliminary; not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 1-115 <ZUR>
C: Keywords: tandem repeat

[illegible]

RESULT 9
T07381
glycine-rich protein Tfm5 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07381
R:Santino, C.G.; Stanford, G.L.; Conner, T.W.
Plant Mol. Biol. 33, 405-416, 1997
A>Title: Developmental and transgenic analysis of two tomato fruit enhanced

A:Accession: T07381
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-207 <S>
A:Cross-references: EMBL:X95262; NID:g1166449; PID:CAA64559.1; PID:g1166450
A:Experimental source: cultivar UC82b; fruit
C:Genetics:
A:Gene: Tfm5
C:Superfamily: hydroxyproline-rich glycoprotein

```

Query Match      6.2%; Score i59.5; DB 2; Length 207;
Best Local Similarity 67.3%; P: 1; Mismatches 12; Indels 3; Gaps 1;
Matches 33; Conservative
QY 240 GSGSGGGSGGSGGGG---QSGSGSGGGSGGGSGGGSGGGSGGGSGG 285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 54 GSGSGGGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGG 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10

Calpain (EC 3.4.22.17) small chain - human
N:Alternate names: calcium-activated neutral proteinase (CANP).
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A26107; A23650
R:R:Miyaake, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 8805-8817, 1986
A:Title: Gene organization of the small subunit of human calcium-activated neutral protein
A:Reference number: A93648; MUID:87066759; PMID:3024120
A:Accession: A26107
A:Molecule type: DNA
A:Residues: 1-268 <MIY>
A:Cross-references: GB:M31502
R:R:Ohno, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 5559, 1986
A:Title: Nucleotide sequence of a cDNA coding for the small subunit of human calcium-de-
A:Reference number: A93631; MUID:86286563; PMID:3016651
A:Accession: A23650
A:Molecule type: mRNA
A:Residues: 1-268 <OHN>
A:Cross-references: EMBL:X04106; NID:g95327; PIDN:CAA27726.1; PID:g95328
C:Genetics:
A:Gene: GDB:CAPN4
A:Cross-references: GDB:119752; OMIM:114170

A; Map position: 19pter-18qter
 A; Introns: 70/2; 81/3; 111/3; 131/1; 152/3; 175/3; 202/1; 241/1; 260/3
 C; Complex: heterodimer of L (large) and S (small) chains
 C; Function:

A:Description: catalyzes the hydrolysis of peptides
 A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and b
 C:Superfamily: calpain small chain; calmodulin repeat homology
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
 F:1-56/Domain: glycine-rich <GLY>
 F:96-127/Domain: calmodulin repeat homology <EF1>
 F:139-171/Domain: calmodulin repeat homology <EF2>
 F:172-201/Domain: calmodulin repeat homology <EF3>
 F:204-236/Domain: calmodulin repeat homology <EF4>
 F:237-268/Domain: calmodulin repeat homology <EF5>

Query Match 6.2%; Score 158.5; DB 1; Length 268;
Best Local Similarity 27.2%; Pred. No. 0.0043;
Matches 66; Conservative 33; Mismatches 101; Indels 43

QY	233	MDLEYKSGGGSGGGSG--SGSGGGQSGG--SGSGGGSGGGSGGGSGGGSGS-----284
Db	4	VNSFLKGGGGGGGGGLGGLGNVLGGLISAGAGGGGGGGGGGGGGGTAMRILG 63
QY	285	-GLRSVKLLSDFDNPRWIGHKHMFNFDWNH-----GKSLDWMYKASDI--331
Db	64	GVTSALSEAAQVNPDPPTTHYSN-IRANESVEVRFRLFAQIAGDDMEVSATELMN 122
QY	332	VINNLGATPEQAKRHRDVEAEFG--GAGMKYGVETDPAYIEGKWKLATDELEYAKN 388
Db	123	ILANKV-----VTRHPDLKTDGFGIDTCRSWAVMDSDTG-----KLGFEF-KYLNW 169
QY	389	EPTLIIRIWDALFDIVDKQNGAITLDEWKAYTKAAGIIQSSEDCETFRVCDIDESGOL 448
Db	170	N----IKRW-QAIYQPDFTDSTGICSSLPAGFAAGFHNLNRYHNIIRYS-DESGNM 224

Qy	449	DVD	451
Dd	225	DEF	227

RESULT 11
T24045

hypothetical protein R08B4.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #t
C/Accession: T24045
R:White, S.

submitted to the EMBL Data Library, November 1995

A;Reference number: Z19834

A;Accession: T24045

A; Status: preliminary; translated from GB/EMBL/DD8.T

A; Molecule type: DNA

A;Residues: 1-1226 <WIL>

A; Cross-References: EMBL:Z68008; PIDN:CAA92000.1; GSPDB:GN00028; CESP:R08B4.1
A; Experimental source: clone R08B4

C;Genetics:

A;Gene: CESP:R08B4.1

A;Map position: X

A; Introns: 53/2; 113/1; 152/3; 204/3; 272/1; 354/1; 389/3; 714/3; 839/3; 877/1; 960/3

Query Match

Query MACCM	0.18;	Score 133.3;	DB 2;	Length 1226;
Best Local Similarity	31.1%;	Pred	NC	0.043.

Best Local Similarity 31.18; Pred: No. 0.043;
Matches 51: Conservative 19: Mismatches

Matches	31;	Conservative	19;	Mismatches	73;	Indels	21;	Gaps	5;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 127 GIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGVSQVLADHYQQNT 186

dbb 749 GLQFEANKVIDQAKQEPDFDFDHNTRNLAKLMTAVQKRGTTFE----SVVAEAFDSWGT 804

187 PIGDGPVLLPDN--HYLSSTQSAISKDPNEKRDHMLVLEFVTAAGITHGMDELYKSGSG 243

db 805 NKENGRTCKIDSOGYNALTYYOS--SSKPPPPPSDFIDIPNDPTI.GGPT-----GSSGGG 855

244 SGGQSGSGSGGQSGSGSGGQSGG-----SGSGGQSGSGSGG 282

[illegible]

RESULT 12

Q0BE31
nuclear antigen EBNA1 - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 22-Oct-1999
C:Accession: C43043; S42440; A03773; S33021
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp ECORI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825

A: Accession: C43043
A: Molecule type: DNA
A: Residues: 1-641 <BMB>
A: Cross-references: EMBL:V01555; NID:g59074; PID:g1334880
A: Experimental source: strain B95-8
R: Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A: Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 R:Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
 proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986
 A:Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a p
 A:Reference number: S42440; MUID:86259739; PMID:3460083

A:Accession: S42440
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-66 <SAM>
A:Cross-references: EMBL:M13941; NID:g330399; PIDN:AAA45889.1; PID:g555157
C:Superfamily: Epstein-Barr virus nuclear antigen
C:Keywords: DNA binding; transcription regulation

[illegible]

RESULT 13
E84537
hypothetical protein At2g16180 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence revision 02-Feb-2001 #text change 02-Feb-2001

C:Accession: E84537
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A84420. MUID:20083487. PMID:10617197

A:Accession: E84537
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1218 <STO>
A:Cross-references: GB:AE002093; NID:g4678213; PIDN:AAD2659.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g16180
A:Map position: 2

[illegible]

403 13HSDU-----GGGFSGGDEGGFSGGDEGGFSGADGEGGFSGGDEGGFSG 437

RESULT 14

I37984

AC3061 12
 Q0BE31
 nucleic antigen EBNA1 - human herpesvirus 4
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 22-Oct-1999
 C:Accession: C43043; S42440; A03773; S33021
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp ECORI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6092825

A: Accession: C43043
A: Molecule type: DNA
A: Residues: 1-641 <BMB>
A: Cross-references: EMBL:V01555; NID:G59074; PID:G1334880
A: Experimental source: strain B95-8
R: Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A: Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 R:Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
 proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986
 A:Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a p
 A:Reference number: S42440; MUID:86259739; PMID:3460083

A:Accession: S42440
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-66 <SAS>
A:Cross-references: EMBL:M13941; NID:g330399; PIDN:AAA45889.1; PID:g555157
C:Superfamily: Epstein-Barr virus nuclear antigen
C:Keywords: DNA binding; transcription regulation

[illegible]

RESULT 13
E84537
hypothetical protein At2g16180 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence revision 02-Feb-2001 #text change 02-Feb-2001

C:Accession: E84537
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A84420. MUID:20083487. PMID:10617197

A:Accession: E84537
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1218 <STO>
A:Cross-references: GB:AE002093; NID:g4678213; PIDN:AAD2659.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g16180
A:Map position: 2

[illegible]

403 13HSDU-----GGGFSGGDEGGFSGGDEGGFSGADGEGGFSGGDEGGFSG 437

Keratin 9, type 1, cytoskeletal - human
 N:Alternate names: Cytokeratin 9; scatter protein 60K chain, placental
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Oct-1999
 C:Accession: I37984; S40307; S77921; S41161; B35494; I37943
 R:Reis, A.; Hennies, H.C.; Langbein, L.; Digweid, M.; Mischke, M.; Sch
 Nature Genet. 6, 174-179, 1994
 A:Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).
 A:Reference number: I37984; MUID:94214498; PMID:7512862
 A:Accession: I37984

A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-622 <RES>
A:Cross-references: EMBL:X75015; NID:q453154; PIDN:CAA52924.1; PID:q453155
R:Langbein, L.; Heid, H.W.; Moll, I.; Franke, W.W.
Differentiation 55, 57-71, 1993
A:Title: Molecular characterization of the body site-specific human epidermal cytoke
A:Reference number: I37943; MUID:94131202; PMID:7507869

A;Accession: S40307
A;Molecule type: mRNA
A;Residues: 1-622 <LAN>
A;Cross-references: EMBL:Z29074; NID:g435475; PIDN:CAA82315.1; PID:g435476
A;Accession: S7921
A;Molecule type: protein

submitted to the EMBL Data Library, December 1993

R:Langbein, L.
A:Residues: 29-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LA>
R:Langbein, L.
A:Residues: 29-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LA>
A:Accession: S41161
A:Molecule type: mRNA
A:Residues: 1-11,'SR',13-622 <LA>

A:Cross-references: EMBL:Z29074; NID:q435475; PIDN:CAA82315.1; PID:q435476
P:Rosen, E. M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.
Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990
A:Title: Human placenta contains an epithelial scatter protein.
A:Reference number: A35494; MUID:90267446; PMID:2140676
A:Accession: B35494
A:Molecule type: protein
A:Residues: 1; X; 450-465 <ROS>

C:Genetics:
A:Gene: GDB: KRT9; EPPK
A:Cross-references: GDB: 303970; OMIM:144200
A:Map position: 17q12-17q21
A:Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2
A:Note: defects in this gene may cause epidermolytic palmoplantar keratoderma
C:Superfamily: cytoskeletal keratin

C: Keywords: coiled coil; intermediate filament
E: 1-153/Domain; head #status predicted <HEA>
F: 134-459/Dcmain: helical rod #status predicted <ROD>
F: 460-622/Dcmain: tail #status predicted <TAI>
Query Match 6.0%; Score 154.5; DB 2; Length 622;
Best Local Similarity 26.6%; Pred. No. 0.022;

	Matches	59: Conservative	31: Mismatches	63: Indels	69: Gaps	11:
QY	107	KTRAEVKFGDGLVNRILKDGIDFKEDGNLGHKLEYNYNHNVYIMADQKNGIKANFK	166			
Db	335	KNRKDIEHQYETQITQIE-----HEVSSSGGEVQSSAKEVTQ	371			
QY	167	IRHNTEGDSVOLADHYQNTPIGDGPVLLPD--NHV-----LSTQSALSKDPNEK	214			

```

Db      372 LRHGQVELEIEIQSLSKAALEKS--LEDTKNRYCGQLOMQTQEISNLEAQITDVRQE 428
QY      215 -----RDHWLL-----BEVTAAGITGHGMDLYKSGSGS---GGQ--SGGS----- 251
Db      429 IEQNQVEYSLLLSIKNRLEKEIETYNLLGEGQDEPFSSGAGKIGLGGSGSGSYGRGS 488
QY      252 -GSGCGGGSGGS-CGQGGSGSGGQS-----GGSGSGCGSQG 285

```

```

Db      489  RGGSGSYGGGGSGGGYGGGSGRGGSGGSGYGGGGSGGGSGG 530

RESULT 15
T31611
hypothetical protein Y50E8A.g - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31611
R:Steward, C.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21047
A:Accession: T31611
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1585 <SWIL>
A:Cross-references: EMBL
A:Experimental source: clone Y50E8A
C:Genetics:
A:Gene: CESP:Y50E8A.g
A:Introns: 25/3; 60/1; 133/2; 217/3; 337/2; 400/1; 746/2

Query Match      6.0%; Score 154; DB 2; Length 1585;
Best Local Similarity 23.3%; Pred. No. 0.072;
Matches 51; Conservative 26; Mismatches 46; Indels 96; Gaps 8;

QY 131 KEDGNTLGHKLEYNHSHNVYIMADKQKNGIKANFKIRNIEDGSVOLADHYQNTPIGD 190
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 660 EEDTQPLRHHHHHHHHH-----OIQLOQLHHLED-----T 691

QY 191 GPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAA-----GITHGMDEL----- 236
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 692 HPEEVEDRH---EEVMLQEDDAEDDLHVVEDTQPLPHRHHHHHQLQHHLEDTHPEST 747

QY 237 -----YKSGSGSGGQSGGSG-----SGG 255
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 748 IWRIRIRGGSGGGYASGGGGGGSGGSGGCRYSYSSAAAPPPPPPPAPAPAPSSGG 807

QY 256 QSGGS-----GSGQGSGSGSGGSGGSG--SGGQSG 285
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 808 YSGGSGGSAAGGGGSGRGIYEGSSGGGGSGGSGYSG 846

Search completed: March 13, 2003, 15:05:45
Job time : 31.1537 secs

```

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:34 ; Search time 15.3871 Seconds
(without alignments)
1285.766 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGEELFTGVFPILVLDG.....LGFWYTMDPACERLYGGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1262	49.3	238	1	GFP_AEQVI	P42212 aequeora vi
2	1032	40.3	196	1	AEQ2_AEQVI	P02592 aequeora vi
3	962	37.6	196	1	AEQ1_AEQVI	P07164 aequeora vi
4	742	29.0	198	1	MYTR_MITCE	P39047 mitrocoma c
5	718	28.0	195	1	OBLT_OBELO	Q27709 obellia long
6	663	25.9	198	1	CLYT_CLYGR	Q08121 clytia greg
7	171.5	6.7	532	1	2IC2_HUMAN	O95409 homo sapien
8	161.5	6.3	450	1	SWP1_ENCCU	Q9XZV1 encephalito
9	161	6.3	115	1	SERI_GALME	O96614 gallieria me
10	158.5	6.2	268	1	CANS_HUMAN	P04632 homo sapien
11	155	6.1	641	1	EBN1_EBV	P03211 Epstein-bar
12	154.5	6.0	622	1	K1C1_HUMAN	P35527 homo sapien
13	151.5	5.9	481	1	LORI_MOUSE	P18165 mus musculu
14	149.5	5.8	204	1	CORA_MESDA	Q07202 medicago sa
15	149	5.8	269	1	CANS_MOUSE	O88456 mus musculu
16	149	5.8	421	1	BR3A_MOUSE	P17208 mus musculu
17	148.5	5.8	627	1	K2C1_MOUSE	O70104 mus musculu
18	148	5.8	1380	1	DDX9_MOUSE	O70103 mus musculu
19	148	5.8	1901	1	Y208_MYCTU	O53553 mycobacteri
20	147.5	5.8	165	1	GRP1_ORISA	P25074 oryza sativ
21	147.5	5.8	434	1	COAA_BPIKE	P03663 bacterioph
22	147	5.7	526	1	K1CJ_BOVIN	P06394 bos taurus
23	147	5.7	569	1	K1CJ_MOUSE	P02535 mus musculu
24	146.5	5.7	266	1	CANS_RABIT	P06813 oryctolagus
25	146	5.7	465	1	GRP2_PHAVU	P10496 phaseolus v
26	145.5	5.7	166	1	K2C5_BOVIN	P04262 bos taurus
27	145	5.7	263	1	CANS_BOVIN	P13135 bos taurus
28	144.5	5.6	266	1	CANS_PIG	P04574 sus scrofa
29	144	5.6	183	1	GRP2_ORISA	P29834 oryza sativ
30	143.5	5.6	266	1	CANS_RAT	Q64537 rattus norv
31	143.5	5.6	434	1	COAA_BPT22	P15415 bacterioph
32	143.5	5.6	593	1	K1C7_HUMAN	P13645 homo sapien
33	143.5	5.6	677	1	SPB7_DICDI	P34643 dictyosteli

RESULT 1

ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RN	(1)			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent protein.";			
RL	Gene 111:229-233(1992).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein.";			
RL	Biochemistry 32:1212-1218(1993).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormoie M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent protein.";			
RL	Science 273:1392-1395(1996).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.			
RX	MEDLINE=98455509; PubMed=9782051;			
RA	Wachter R.M., Ellsiger M.A., Kallio K., Hanson G.T., Remington S.J.;			
RT	"Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein.";			
RL	Structure 6:1267-1277(1998).			

ALIGNMENTS

34	141.5	5.5	759	1	GSPD_XANCP	P29041 xanthomonas
35	141.5	5.5	1250	1	TP3A_DROME	Q9ng98 drosophila
36	141	5.5	1726	1	MSPI_PLAFC	P04934 plasmodium
37	141	5.5	1726	1	MSPI_PLAFC	P50495 plasmodium
38	140	5.5	322	1	PUR_HUMAN	Q05777 homo sapien
39	140	5.5	333	1	SIX3_MOUSE	Q62233 mus musculu
40	139.5	5.4	316	1	LORI_HUMAN	P23490 homo sapien
41	139.5	5.4	384	1	GRP1_PETHY	P09789 petunia hyb
42	139	5.4	424	1	COAA_BPFD	P03661 bacterioph
43	139	5.4	424	1	COAA_BPFD	P03662 bacterioph
44	138.5	5.4	148	1	CLM4_MOUSE	Q9jm83 mus musculu
45	138.5	5.4	734	1	YKR2_CABEL	P34308 caenorhabdi

[7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=93238303; PubMed=10220315;
RA Elsiger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
RT "Structural and spectral response of green fluorescent protein
variants to changes in pH."
RL Biochemistry 38:5296-5301(1999).
CC -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
CC BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN
CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
CC RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN Aequorin.
CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
CC ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
CC -!- SURUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: PHOTOCYTES.
CC -!- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
CC chimeric proteins of GFP linked to other proteins where it
CC functions as a fluorescent protein tag. GFP tolerates N- and C-
CC terminal fusion to a broad variety of proteins. It has been
CC expressed in bacteria, yeast, slime mold, plants, drosophila,
CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
CC marker in living cells, it allows for a wide range of applications
CC where it may function as a cell lineage tracer, reporter of gene
CC expression, or as a measure of protein-protein interactions.
CC -!- DATABASE: NAME-Protein Spotlight;
CC NOTE-Issue 11 of June 2001;
CC WWW="http://www.expasy.org/spotlight/articles/splt011.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62654; AAA27722.1; -
DR EMBL: M62653; AAA27721.1; -
DR EMBL: L29345; AAA58246.1; -
DR PDB: 1GFL; 11-JAN-97.
DR PDB: 1EMA; 08-NOV-96.
DR PDB: 1EMB; 16-JUN-97.
DR PDB: 1EMC; 20-AUG-97.
DR PDB: 1EMG; 20-AUG-97.
DR PDB: 1EMD; 20-AUG-97.
DR PDB: 1EME; 20-AUG-97.
DR PDB: 1EMF; 20-AUG-97.
DR PDB: 1EMG; 12-MAY-99.
DR PDB: 1EMK; 20-AUG-97.
DR PDB: 1EMI; 20-AUG-97.
DR PDB: 1EMW; 20-AUG-97.
DR PDB: 2EMN; 20-AUG-97.
DR PDB: 2EMO; 20-AUG-97.
DR PDB: 1BFP; 07-JUL-97.
DR PDB: 1YFP; 28-OCT-98.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFPLORESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
KW Luminescence; 3D-structure.
FT SITE 65 67
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
SQ SEQUENCE 238 AA; 26886 MW; EA5A6F21FBFB6B05 CRC64;

Query Match 49.3%; Score 1262; DB 1; Length 238;
Best Local Similarity 98.7%; Pred. No. 3e-69;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYGKLTLLKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLDGVDVNGHKFSVSGEGDGYGKLTLLKFKICTTGKLPVWPPTL 60
QY 61 VTTUTYGVQCFSPYDPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
DB 61 VTTUTYGVQCFSPYDPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHVYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNHNHVYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPTGDPVLLPDNHYLSTQSLKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPTGDPVLLPDNHYLSTQSLKDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
RESULT 2
AEQ2_AEQVI STANDARD; PRT; 196 AA.
ID AEQ2_AEQVI
AC P02592;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aequorin 2 precursor.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85216460; PubMed=3858813;
RA Inouye S., Noguchi M., Sakaki Y., Takagi Y., Miyata T., Iwanaga S.,
RA Miyata T., Tsuji F.I.;
RT "Cloning and sequence analysis of cDNA for the luminescent protein
RT aequorin."
RL proc. Natl. Acad. Sci. U.S.A. 82:3154-3158(1985).
RN [2]
RP SEQUENCE OF 9-185 FROM N.A. (AEQUORIN 2 AND 3).
RX MEDLINE=87185437; PubMed=2882777;
RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;
RT "Sequence comparisons of complementary DNAs encoding aequorin
RT isotypes."
RL Biochemistry 26:1326-1332(1987).
RN [3]
RP SEQUENCE OF 8-196.
RX MEDLINE=86077721; PubMed=2866797;
RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,
RA Cormier M.J., Vanaman T.C.;
RT "Amino acid sequence of the calcium-dependent photoprotein aequorin."
RL Biochemistry 24:6762-6771(1985).
RN [4]
RP MUTAGENESIS.
RA Tsuji F.I., Inouye S., Goto T., Sakaki Y.;
RT "Site-specific mutagenesis of the calcium-binding photoprotein
RT aequorin."
RL proc. Natl. Acad. Sci. U.S.A. 83:8107-8111(1986).
RN [5]
RP MUTAGENESIS OF PRO-196.
RX MEDLINE=9211761; PubMed=1765170;
RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;
RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-
RT binding photoprotein, aequorin."
RL FEBS Lett. 295:63-66(1991).
RN [6]
RP DISULFIDE BOND.
RX MEDLINE=94009705; PubMed=8405461;
RA Ohmiya Y., Kuroso S., Ohashi M., Fagan T.F., Tsuji F.I.;
RT "Mass spectrometric evidence for a disulfide bond in aequorin


```
DR Pfam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EPH; 2.
DR PROSITE: PS00018; EF_HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 7
FT CHAIN 8 196
FT SITE 47 57
FT SITE 62 72
FT SITE 107 117
FT CA_BIND 31 42
FT DOMAIN 72 88
FT CA_BIND 124 135
FT CA_BIND 160 171
FT DISULFID 152 159
SQ SEQUENCE 196 AA; 22514 MW; 9AA5B636288A5B8F CRC64;

Query Match 37.6%; Score 962; DB 1; Length 196;
Best Local Similarity 90.5%; Pred.No. 2.3e-51;
Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPRIWGRHKHMFNFDVNHNGKISLDEMYKASDIVINNLGATPEQAKRHK 347
DB 7 SVKLTDFDNPRIWGRHKHMFNFDVNHNGKISLDEMYKASDIVINNLGATPEQAKRHK 66
QY 348 DAVEAFEGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 407
DB 67 DAVEAFEGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKD 126
QY 408 QNGAITLDWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMOPA 467
DB 127 QNGAISLDWKAYTKSDGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMOPA 186
QY 468 CEKLYGGAVP 477
DB 187 CEKLYGGAVP 196

RESULT 4
MYTR_MITCE
ID MYTR_MITCE STANDARD; PRT; 198 AA.
AC P39047;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitrocomin precursor.
GN M17.
OS Mitrocoma cellularia (Halistauro mitrocoma).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
OC Mitrocomidae; Mitrocoma.
OX NCBI_TaxID=31874;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039781; PubMed=8224198;
RA Fagan T.F., Ohmura Y., Blinks J.R., Inouye S., Tsuji F.I.;
RT "Cloning, expression and sequence analysis of cDNA for the Ca(2+)-
RT binding photoprotein, mitrocomin.";
RL FEBS Lett. 333:301-305(1993).
CC -1- FUNCTION: CA(2+)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COLENTERAZINE INTO COLENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin family.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


or send an email to license@isb-sib.ch.

```

CC EMBL; U07128; AAA67708.1; -.
CC HSSP; P02592; 1EJ3.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; efhand; 3.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00054; EFh; 2.
CC PROSITE; PS00018; EF-HAND; 3.
CC Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 195
FT CHAIN 7 195
FT CA_BIND 30 41
FT DOMAIN 71 87
FT CA_BIND 123 134
FT CA_BIND 159 170
FT CA_BIND 151 158
FT DISULFID 151 158
SQ SEQUENCE 195 AA; 22226 MW; 5D002270B73D3663 CRC64;

Query Match 28.0%; Score 718; DB 1; Length 195;
Best Local Similarity 68.4%; Pred. No. 9.8e-37;
Matches 130; Conservative 23; Mismatches 37; Indels 0; Gaps 0;

Qy 288 SVKLTSDPNRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
Db 6 AVKLKTDSDPNRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 65
Qy 348 DAVEAFFGAGMKYGVETDPAYTEGWKKLATDELYAKNEPTLIRIWGALFDIVDKD 407
Db 66 VCVEAFFGCGMEYKETAFFQFDGKQKQATSELUKWARNEPTLIRIWGALFDIVDKD 125
Qy 408 ONGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESSGOLDVDVDMTRHGLGFWYTMDDPA 467
Db 126 GSGTITLDEWKAYKISGISPSQEDCEATFRHCDLNSGDLVDVDMTRHGLGFWYTMDDPA 185
Qy 468 CEKLYGGAVP 477
Db 186 ADGLYNGVVP 195

RESULT 6
CLYT_CLYGR STANDARD; PRT; 198 AA.
AC Q08121;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clytin precursor (Phalidid).
OS Clytia gregaria (Phalidium gregarium).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae; Campanularidae; Clytia.
OX NCBI_TaxID=27801;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93138101; PubMed=8422928;
RA Inouye S., Tsuji F.I.;
RT "Cloning and sequence analysis of cDNA for the Ca(2+)-activated photoprotein, clytin."
RL FEBS Lett. 315:343-346(1993).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE, COELENTAZINE INTO COELENTERAMIDE AND CO(2) WITH THE CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

or send an email to license@isb-sib.ch.

```

CC EMBL; L13247; AAA28293.1; -.
CC EMBL; X70221; CAA49754.1; -.
CC PIR; S28860; S28860.
CC HSSP; P02592; 1EJ3.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; efhand; 3.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00054; EFh; 2.
CC PROSITE; PS00018; EF-HAND; 3.
CC Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 198
FT CHAIN 10 198
FT CA_BIND 33 44
FT DOMAIN 74 90
FT CA_BIND 126 137
FT CA_BIND 162 173
FT DISULFID 148 161
SQ SEQUENCE 198 AA; 22541 MW; 23F1E399667F9059 CRC64;

Query Match 25.9%; Score 663; DB 1; Length 198;
Best Local Similarity 61.6%; Pred. No. 2e-33;
Matches 117; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

Qy 288 SVKLTSDPNRWIGRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
Db 9 AVKLKRPEDNPKWNRHKHMFNFDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 68
Qy 348 DAVEAFFGAGMKYGVETDPAYTEGWKKLATDELYAKNEPTLIRIWGALFDIVDKD 407
Db 69 DAVEAFFGAGMKYGVETDPAYTEGWKKLATDELYAKNEPTLIRIWGALFDIVDKD 128
Qy 408 ONGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESSGOLDVDVDMTRHGLGFWYTMDDPA 467
Db 129 GSGTITLDEWKAYKISGISPSQEDCEATFRHCDLNSGDLVDVDMTRHGLGFWYTMDDPA 188
Qy 468 CEKLYGGAVP 477
Db 189 ADGLYNGVVP 198

RESULT 7
ZIC2_HUMAN STANDARD; PRT; 532 AA.
ID ZIC2_HUMAN STANDARD; PRT; 532 AA.
AC O95409; Q9H309;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE zinc finger protein ZIC2 (zinc finger protein of the cerebellum 2).
GN ZIC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. AND VARIANT HPES POLY-ALA INSERTION.
RX MEDLINE=98442655; PubMed=9771712;
RA Brown S.A., Warburton D., Brown L.Y., Yu C.Y., Roeder E.R., Stengel-Rutkowski S., Hennekam R.C., Muenke M.;
RT "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired."
RL Nat. Genet. 20:180-183(1998).
RN [2]
SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA MEDLINE=20556339; PubMed=10984499;
RA Yang Y., Hwang C.K., Junn E., Lee G., Mouradian M.M.;
RT "ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine receptor gene."
RL J. Biol. Chem. 275:38863-38869(2000).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: DEFECTS IN ZIC2 ARE A CAUSE OF HOLOPROSENCEPHALY TYPE 5

```

```

CC (HPE5). HPE5 IS A STRUCTURAL ANOMALY OF THE BRAIN.
CC -!- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF104902; AAC96325.1; -
DR EMBL; AF193855; AAG28409.1; -
DR HSSP; P08047; ISP2.
DR TRANSFAC; T04237; -
DR Genew; HGNC:12873; ZIC2.
DR MIM; 603073; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; 2f-C2H2; 5.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Disease mutation; Holoprosencephaly.
FT DOMAIN 20 23 POLY-HIS.
FT DOMAIN 25 33 POLY-ALA.
FT DOMAIN 89 97 POLY-ALA.
FT DOMAIN 226 230 POLY-ALA.
FT DOMAIN 231 239 POLY-HIS.
FT DOMAIN 300 415 ZINC FINGERS.
FT ZN_FING 300 327 C2H2-TYPE (ATYPICAL).
FT ZN_FING 333 357 C2H2-TYPE.
FT ZN_FING 363 387 C2H2-TYPE.
FT ZN_FING 393 415 C2H2-TYPE.
FT DOMAIN 456 470 POLY-ALA.
FT DOMAIN 490 508 POLY-GLY.
FT VARIANT 470 470 A -> AAAAAAAAAA (IN HPE5).
FT
FT CONFLICT 124 128 RGFED -> ARLPGT (IN REF. 1).
FT SEQUENCE 532 AA; 55006 MW; BA3E6455DAF97EAC CRC64;

Query Match 6.7%; Score 171.5; DB 1; Length 532;
Best Local Similarity 31.0%; Pred. No. 0.0017;
Matches 54; Conservative 24; Mismatches 73; Indels 23; Gaps 8;

QY 124 ELKGIKEDKGNLGHKLEYNHNYI--MADK---OKNGIKANFKIRHNIEDGSVQ- 177
Db 366 EFGCDRRFANSDDRKHKHMHVHTSDKPYLCKMKDKSYTHPSSLRKHMKVHSSPQSSSS 425
QY 178 --LADHYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMLVLEFVTAGITHGMDE 235
Db 426 PAASSGYESTPPG---LVSPS---AEPOSSSNLSP-----AAAAAAAAAAAVSA 473
QY 236 LYKSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 288
Db 474 VHRGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 525

RESULT 8
SWP1_ENCCU STANDARD; PRT; 450 AA.
AC Q9X2V1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Spore wall protein 1 precursor.
GN SWP1.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Bohne W., Ferguson D.J.P., Kohler K., Gross U.;
RT "Molecular characterisation of a developmentally expressed spore wall
RL protein from the human microsporidian Encephalitozoon cuniculi.";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPORE WALL COMPONENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ133745; CAB39735.1; -
KW Sporulation; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 450 SPORE WALL PROTEIN 1.
FT DOMAIN 66 74 POLY-SER.
FT DOMAIN 79 82 POLY-ARG.
FT DOMAIN 303 339 THR-RICH.
FT DOMAIN 355 450 GLY/SER-RICH.
SQ SEQUENCE 450 AA; 45873 MW; 5E7071A3E3A6DF60 CRC64;

Query Match 6.3%; Score 161.5; DB 1; Length 450;
Best Local Similarity 55.4%; Pred. No. 0.0057;
Matches 41; Conservative 1; Mismatches 5; Indels 27; Gaps 4;

QY 239 SGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 274
Db 373 SDGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 432
QY 275 SG---SGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 285
Db 433 SGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 446

RESULT 9
SER1_GALME STANDARD; PRT; 115 AA.
AC O96614;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Sericin-1 (Silk gum protein 1) (Fragment).
GN SER1 OR SER-1.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Pyraloidea; Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Middle silk gland;
RA Zurevec M., Sehnael F., Scheller K., Kumaran A.K.;
RT "Silk gland specific cDNAs from Galleria mellonella L.";
Submitted (BIO-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC SECTION OF SILK GLANDS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF095241; AAC79078.1; -

```


[8]
 RP VARIANT EPPK VAL-159.
 RX MEDLINE=97348961; PubMed=9204965;
 RA Endo H., Hatamochi A., Shinkai H.;
 RT "A novel mutation of a leucine residue in coil 1A of keratin 9 in
 epidermolytic palmoplantar keratoderma.";
 RL J. Invest. Dermatol. 109:113-115(1997).
 RN [9]
 RP VARIANTS EPPK THR-156; VAL-156 AND GLN-162.
 RX MEDLINE=99072662; PubMed=9856842;
 RA Covello S.P., Irvine A.D., McKenna K.E., Munro C.S., Nevlin N.C.,
 RA Smith F.J.D., Utto J., McLean W.H.I.;
 RT "Mutations in keratin 9 in kindreds with epidermolytic palmoplantar
 keratoderma and epidemiology in Northern Ireland.";
 RL J. Invest. Dermatol. 111:1207-1209(1998).
 CC -1- FUNCTION: MAY SERVE AN IMPORTANT SPECIAL FUNCTION EITHER IN THE
 MATURE PALMAR AND PLANTAR SKIN TISSUE OR IN THE MORPHOGENIC
 PROGRAM OF THE FORMATION OF THESE TISSUE.
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE TERMINALLY DIFFERENTIATED
 EPIDERMIS OF PALMS AND SOLES.
 CC -1- DISEASE: DEFECTS IN KRT9 ARE A CAUSE OF EPIDERMOLYTIC PALMOPLANTAR
 KERATODERMA (EPPK), AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY
 DIFFUSE THICKENING OF THE EPIDERMIS ON THE ENTIRE SURFACE OF
 PALMS AND SOLES SHARPLY BORDERED WITH ERYTHEMATOUS MARGINS.
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
 MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
 (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE A 60 kDa CHAIN OF
 PLACENTAL SCATTER PROTEIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

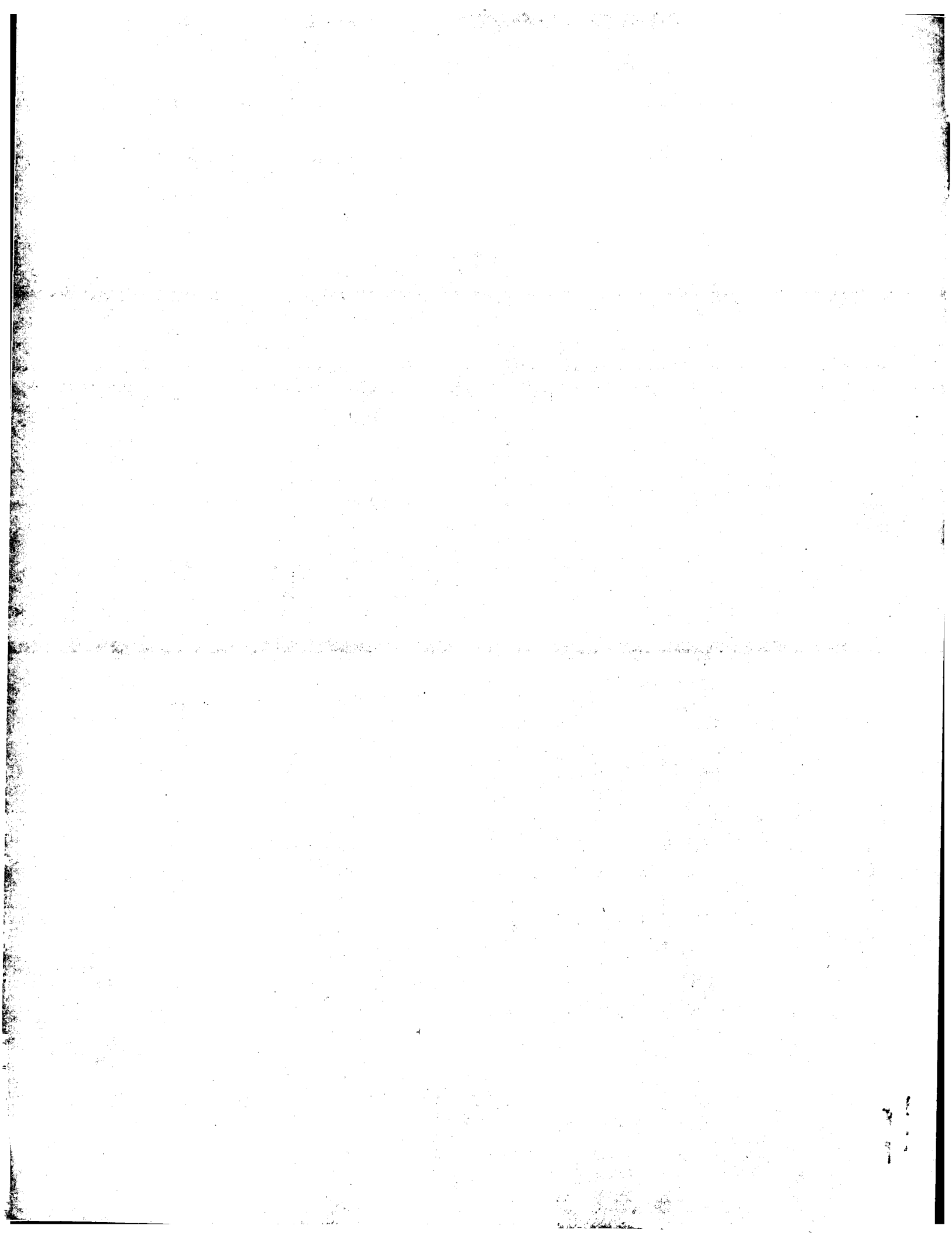
 DR EMBL; X75015; CAA52924.1; -;
 DR EMBL; Z29074; CAA82315.1; -;
 DR EMBL; S69510; AAC60619.1; -;
 DR PIR; B35494; B35494.
 DR Genew; HGNC:6447; KRT9.
 DR MIN; 144200; -;
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin_I.
 DR Pfam; PF00038; Filament; I.
 DR PRINTS; PR01248; TYPEIKERATIN.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Keratin; Disease mutation.
 FT DOMAIN 1 151
 FT HEAD.
 FT DOMAIN 152 460
 FT ROD.
 FT DOMAIN 461 622
 FT TAIL.
 FT DOMAIN 152 187
 FT COIL 1A.
 FT DOMAIN 188 206
 FT LINKER 1.
 FT COIL 1B.
 FT DOMAIN 207 298
 FT LINKER 12.
 FT DOMAIN 299 321
 FT LINKER 2.
 FT DOMAIN 322 460
 FT COIL 2.
 FT DOMAIN 14 20
 FT POLY-GLY.
 FT DOMAIN 156 156
 FT /FTid=VAR_010499.
 FT M -> T (IN EPPK).
 FT /FTid=VAR_010499.
 FT M -> V (IN EPPK).
 FT L -> V (IN EPPK).
 FT N -> K (IN EPPK).
 FT /FTid=VAR_010501.
 FT N -> S (IN EPPK).
 FT /FTid=VAR_003822.
 FT N -> S (IN EPPK).
 FT /FTid=VAR_010502.
 FT N -> Y (IN EPPK).
 FT /FTid=VAR_010503.

FT VARIANT 162 162 R -> Q (IN EPPK).
 FT /FTid=VAR_003823.
 FT VARIANT 162 162 R -> W (IN EPPK).
 FT /FTid=VAR_003824.
 FT VARIANT 167 167 L -> S (IN EPPK).
 FT /FTid=VAR_003825.
 FT VARIANT 171 171 Q -> P (IN EPPK).
 FT /FTid=VAR_010504.
 FT CONFLICT 12 12 T -> SR (IN REF. 1; CAA82315).
 SQ SEQUENCE 622 AA; 61987 MW; 898C3825D4B5ED94 CRC64;
 Query Match 6.0%; Score 154.5; DB 1; Length 622;
 Best Local Similarity 26.6%; Pred. No. 0.021;
 Matches 59; Conservative 31; Mismatches 63; Indels 69; Gaps 11;
 QY 107 KTRAEVKEGDTLVNRIELKGDIFKEDGNILGHKLEYNVNSHNVIYIMADKOKNGIKANFK 166
 DB 335 KNRKDIEHQYETQITQIE-----HEVSSSGOEVOSSAKEVTQ 371
 QY 167 IRHNIEDGVSQADHYQOQTPIGDGPVLLPD--NHV-----LSTQSALSADPNK 214
 DB 372 LRHGVOELLEIELOLSKKALEKS--LEDTKNRYCGOLOMIOEQISNLEAQITDVQE 428
 QY 215 -----RDHVVLL-----EFVTAAGTHGMDELYKSGSGS---GGQ--SGGS----- 251
 DB 429 IECQNOEYSLLSIKMRLEKETIYHNLGGQEDFESSGAGKIGLGGSGSGSGYGRGS 488
 QY 252 --GGSGSGSGSGS--GGSGSGSGSGS-----GGSGSGSGSG 285
 DB 489 RGGSGSGSGSGYGGSGSGSGSGSGSGSGSGSGSGSGSGSG 530
 RESULT 13
 LORI_MOUSE
 ID LORI_MOUSE STANDARD; PRT; 481 AA.
 AC P18165;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Loricrin.
 GN LOR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90275605; PubMed=2190691;
 RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
 RA Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
 RA Yuspa S.H., Roop D.R.;
 RT "Identification of a major keratinocyte cell envelope protein,
 loricrin";
 RL Cell 61:1103-1112(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RX MEDLINE=95256248; PubMed=7738016;
 RA DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
 RA Roop D.R.;
 RT "The proximal promoter of the mouse loricrin gene contains a
 functional AP-1 element and directs keratinocyte-specific but not
 differentiation-specific expression.";
 RL J. Biol. Chem. 270:10792-10799(1995).
 CC -1- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
 CC -1- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
 GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M34398; AAA39444.1; -.  
DR EMBL; U09189; AAA82152.1; -.  
DR PIR; A35628; A35628.  
DR HSSP; P02876; 9WGA.  
DR MGD; MGI:96816; Lor.  
KW Keratinization.  
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;  
  
Query Match 5.9%; Score 151.5; DB 1; Length 481;  
Best Local Similarity 45.2%; Pred.No. 0.024; 33; Indels 11; Gaps 3;  
Matches 38; Conservative 2; Mismatches 3;  
  
QY 228 GITHGMDELYKSGGS-SGGQSGGS---SGQSGGS-----SGQSGGS 276  
DB 395 GSGGGGGYSGGGGCGGSGGSGGCGGSGGSGGCGGSGGSGGSGGSGGSGGSS 454  
  
QY 277 GSGSGGQSLRSVKLTSTDFDNPW 300  
DB 455 GGGSGGKGVPVCHQTOQKQAPTW 478  
  
RESULT 14  
COR_MEDSA STANDARD; PRT; 204 AA.  
AC Q07202;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Cold and drought-regulated protein CORA.  
GN Medicago sativa (Alfalfa).  
OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
OX NCBI_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Apica;  
RX MEDLINE=94143496; PubMed=8310076;  
RA Laberge S., Castonguay Y., Vezina L.-P.;  
RT "New cold- and drought-regulated gene from Medicago sativa.";  
RL Plant Physiol. 101:1411-1412(1993).  
CC -!- FUNCTION: MAY BE INVOLVED IN RESISTANCE OF THE PLANT TO  
ENVIRONMENTAL STRESS.  
CC -!- INDUCTION: BY COLD, ABSCISIC ACID (ABA) AND DROUGHT STRESS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; L03708; AAA99833.1; -.  
DR Multigene family; Repeat.  
KW 7 X 6 AA REPEATS OF Y-N-H-G-G-G.  
FT DOMAIN 54 176  
FT REPEAT 54 59  
FT REPEAT 65 70  
FT REPEAT 71 76  
FT REPEAT 78 83  
FT REPEAT 85 90  
FT REPEAT 164 169  
FT REPEAT 171 176  
FT REPEAT 192 192  
FT DOMAIN 98 100  
FT REPEAT 98 100  
FT REPEAT 101 103  
FT REPEAT 112 114  
FT REPEAT 115 117  
FT REPEAT 126 128  
FT REPEAT 128 128  
  
us-09-863-901-5-rsp
```

```
FT REPEAT 129 131  
FT REPEAT 178 180  
FT REPEAT 181 183  
FT REPEAT 184 186  
FT REPEAT 187 189  
FT REPEAT 190 192  
SQ SEQUENCE 204 AA; 19599 MW; 8A2C082359FCC17F CRC64;  
  
Query Match 5.8%; Score 149.5; DB 1; Length 204;  
Best Local Similarity 35.4%; Pred.No. 0.012;  
Matches 35; Conservative 11; Mismatches 44; Indels 9; Gaps 3;  
  
QY 228 GITHGMDELYKSGGS---GSGQSGGS-SGGQSGGS-SGGQSGGS-SGGQSGGS 283  
DB 70 GYNHGGGGYHNGGGYHNGGGGGHGGGGHGGGGYHNGGGGGHGGGGHGGGGHGGH 129  
  
QY 284 SGLRSVKL-----TSDFDNPWIRGRHKHMFNFDVNHNG 317  
DB 130 GGAESVAVOTEKTEKNEVDKARYGGGSNYNDGRGGYNHGG 168  
  
RESULT 15  
CANS_MOUSE STANDARD; PRT; 269 AA.  
ID CANS_MOUSE  
AC O88456; Q8VEK4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)  
DE (Calcium-activated neutral proteinase) (CANP).  
GN CAPNS1 OR CAPN4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20285452; PubMed=10825211;  
RA Arthur J.S.C., Elce J.S., Hegadorn C., Williams K., Greer P.A.;  
RT "Disruption of the murine calpain small subunit gene, Capn4; calpain  
RT is essential for embryonic development but not for cell growth and  
RT division.";  
RL Mol. Cell. Biol. 20:4474-4481(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which  
CC catalyze limited proteolysis of substrates involved in  
CC cytoskeletal remodeling and signal transduction.  
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small  
CC (regulatory) subunit.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma  
CC membrane upon Ca++ binding (By similarity).  
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AF058298; AAC97194.1; -.  
DR EMBL; BC018352; AAH18352.1; -.  
DR HSSP; P04632; 1KFU.  
DR MGD; MGI:88266; Capns1.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; ehand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR PROSITE; PS00018; EF_HAND; 2.  
KW Calcium-binding; Repeat.
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 15:00:59 ; Search time 47.0664 Seconds
(without alignments)
2088.208 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2560
Sequence: 1 MSKGELFTGVVPLVELDG.....LGFWYTMDPAGEKLYGGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1256	49.1	238	5 Q93125	Q93125 aequorea vi
2	1254	49.0	238	5 Q27903	Q27903 unidentified
3	1231	48.1	758	4 Q96JQ1	Q96JQ1 homo sapien
4	1220	47.7	238	5 Q17105	Q17105 aequorea vi
5	1205	47.1	238	5 Q17106	Q17106 aequorea vi
6	1100	43.0	238	5 Q8WTC6	Q8WTC6 aequorea ma
7	1096	42.8	238	5 Q8WPC9	Q8WPC9 aequorea ma
8	1093	42.7	238	5 Q8WTC4	Q8WTC4 aequorea ma
9	1091	42.6	238	5 Q8WTD0	Q8WTD0 aequorea ma
10	1090	42.6	238	5 Q8WTC9	Q8WTC9 aequorea ma
11	1090	42.6	238	5 Q8WTC8	Q8WTC8 aequorea ma
12	1087	42.5	238	5 Q8WTC7	Q8WTC7 aequorea ma
13	1086	42.4	238	5 Q8WTC5	Q8WTC5 aequorea ma
14	915	35.7	195	5 Q8WQY8	Q8WQY8 aequorea pa
15	899	35.1	195	5 Q8WQY7	Q8WQY7 aequorea ma
16	705	27.5	195	5 Q8T620	Q8T620 obelia geni

17	253.5	9.9	225	5	Q95UA7	Q95UA7 montastraea
18	246	9.6	225	5	Q963F5	Q963F5 montastraea
19	240	9.4	236	5	Q8T6U0	Q8T6U0 dendronephth
20	233.5	9.1	266	5	Q9U6Y3	Q9U6Y3 clavularia
21	231.5	9.0	225	5	Q8T5F1	Q8T5F1 montastraea
22	210	8.2	227	5	Q962P9	Q962P9 montastraea
23	208	8.1	229	5	Q9U6Y6	Q9U6Y6 anemonia ma
24	207.5	8.1	234	5	Q8T5F2	Q8T5F2 montastraea
25	205.5	8.1	235	5	Q8T5F0	Q8T5F0 scolymia cu
26	205.5	8.0	232	5	Q9GP15	Q9GP15 anemonia su
27	204.5	8.0	225	5	Q9U6Y8	Q9U6Y8 discosoma s
28	204	8.0	227	5	Q95VT0	Q95VT0 montastraea
29	203.5	7.9	238	5	Q9BLY9	Q9BLY9 renilla mue
30	201.5	7.9	232	5	Q9GZ28	Q9GZ28 anemonia su
31	201	7.9	221	5	Q95P04	Q95P04 gonolopora t
32	193.5	7.8	225	5	Q8T6T9	Q8T6T9 heteractis
33	194	7.6	232	5	Q9U6Y7	Q9U6Y7 discosoma s
34	192	7.5	230	5	Q9GTJ7	Q9GTJ7 discosoma s
35	187.5	7.3	227	5	Q95W85	Q95W85 heteractis
36	185.5	7.3	228	5	Q9GP16	Q9GP16 anemonia su
37	182.5	7.1	238	5	Q9BLZ0	Q9BLZ0 ptilosarcus
38	181.5	7.1	233	5	Q963I9	Q963I9 renilla ren
39	179.5	7.0	227	5	Q95W86	Q95W86 condylactis
40	179.5	7.0	231	5	Q9U6Y5	Q9U6Y5 zoanthus sp
41	175	6.8	229	5	Q8T5E7	Q8T5E7 condylactis
42	173.5	6.8	227	5	Q95W11	Q95W11 condylactis
43	170	6.6	302	10	Q9SL09	Q9SL09 arabidopsis
44	170	6.6	738	5	O02402	O02402 pinctada fu
45	169	6.6	125	5	Q964C1	Q964C1 encephalito

ALIGNMENTS

RESULT 1

Q93125	ID	Q93125	PRELIMINARY;	PRT;	238 AA.
AC	Q93125;				
DT	01-FEB-1997 (Tremblrel. 02, Created)				
DT	01-FEB-1997 (Tremblrel. 02, Last sequence update)				
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)				
DE	Green fluorescent protein mutant 3.				
GN	GFP.				
OS	Aequorea victoria (Jellyfish).				
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;				
OC	Aequoreidae; Aequorea.				
OX	NCBI_TaxID=6100;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96305137; PubMed=8707053;				
RA	Cornack B.P., Valdivia R.H., Falkow S.;				
RT	"FACS-optimized mutants of the green fluorescent protein (GFP).";				
RL	Gene 173:33-38(1996).				
RP	[2]				
RP	SEQUENCE FROM N.A.				
RA	Cornack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,				
RA	Brown A.J.P.;				
RT	"Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene				
RT	expression in Candida albicans.";				
RL	Microbiology 0:0-0(1996).				
DR	EMBL; U73901; AAB18957.1; -.				
DR	HSSP; P42212; 1BPP.				
DR	InterPro; IPR000786; Green_fl_protein.				
DR	Pfam; PF01353; GFP; 1.				
DR	PRINTS; PRO1229; GFLUORESCENT.				
DR	ProDom; PD013756; Green_fl_protein; 1.				
SQ	SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;				
Query Match					
Best Local Similarity 49.1%; Score 1256; DB 5; Length 238;					
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;					
QY	1	MSKGELFTGVVPLVELDGVNKHKFSVSGEGEGDATYKGLTKFKICTTGGKLPVWPPTL 60			

```

|||||
Db 1 MSKGEELFTGVVPLVLDGDVNGHKFVSVEGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
Db 61 VTFEGYGVQCFARYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITHGMDELYK 238

RESULT 2
Q27903 ID Q27903 PRELIMINARY; PRT; 238 AA.
AC Q27903;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9729982; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
  protein by modification of its codon usage.";
RL Plant Mol. Biol. 33:989-999(1997).
DR EMBL; X96418; CAA65278.1; -.
DR HSSP; P42212; IGFL.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match 49.0%; Score 1254; DB 5; Length 238;
Best Local Similarity 97.9%; Pred. NO. 8.3e-82;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGDVNGHKFVSVEGEGDATYKGLTLKFICTTGKLPVWPPTL 60
Db 1 MGKGEELFTGVVPLVLDGDVNGHKFVSVEGEGDATYKGLTLKFICTTGKLPVWPPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
Db 61 VTFEGYGVQCFARYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITHGMDELYK 238

RESULT 3
Q96J01 ID Q96J01 PRELIMINARY; PRT; 758 AA.
AC Q96J01;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Raichu404X.
GN RAICHU404X.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RX MEDLINE=21322811; PubMed=11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
  Miyawaki A., Watsuda M.;
RT "Spacio-temporal images of Growth Factor-induced Activation of Ras and
  Rapi.1.";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_transfmg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; ras; 1.
DR Pfam; PF02196; RBD; 1.
DR ProDom; PD013756; Green_fl_protein; 2.
DR TIGRFAMS; TIGR00231; small_gtp; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
SQ SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;

Query Match 48.1%; Score 1231; DB 4; Length 758;
Best Local Similarity 94.3%; Pred. NO. 1.9e-79;
Matches 233; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGDVNGHKFVSVEGEGDATYKGLTLKFICTTGKLPVWPPTL 60
Db 499 VSKGEELFTGVVPLVLDGDVNGHKFVSVEGEGDATYKGLTLKFICTTGKLPVWPPTL 558
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 120
Db 559 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYYVQERTIFFKDDGNYKTRAEVKEFGDTLV 618
QY 121 NRLELKGIDFKEDGNILGHKLEYNINSHNYYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 619 NRLELKGIDFKEDGNILGHKLEYNINSHNYYITADKQKNGIKANFKIRHNIEDGSVOLAD 678
QY 181 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITHGMDELYKSG 240
Db 679 HYQONTPIGDPVLLPDNHNHLSQTSALSADPNKRDHVMVLEFVTAAGITLGNDELGRSR 738
QY 241 GSGSGGQ 247
Db 739 KMSKDGK 745

RESULT 4
Q17105 ID Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR HSSP; P42212; IGFL.
DR InterPro; IPR000786; Green_fl_protein.

```

```
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 47.7%; Score 1220; DB 5; Length 238;
Best Local Similarity 94.5%; Pred. No. 2.2e-79;
Matches 225; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||
DB 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||

QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||
DB 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||

RESULT 5
Q17106
ID Q17106 PRELIMINARY; PRT; 238 AA.
AC Q17106;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein (fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83960; CAA58790.1; -.
DR HSP: P42212; 1BPP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 47.1%; Score 1205; DB 5; Length 238;
Best Local Similarity 93.7%; Pred. No. 2.6e-78;
Matches 223; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||
DB 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||
DB 61 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||

QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||
DB 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||
```

```
RESULT 6
Q8WTC6
ID Q8WTC6 PRELIMINARY; PRT; 238 AA.
AC Q8WTC6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RL "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435431; AAL33916.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27015 MW; 6B8FD75E88926903 CRC64;

Query Match 43.0%; Score 1100; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 8.1e-71;
Matches 197; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

QY 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||
DB 1 MSKGELEFTGVVPIILVELDGVNGHKFSVSGEGGDATYKGLTLKFICTTGKLPVPWPTL 60
   |||||

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||
DB 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNKYKTRAEVKFEGDTLV 120
   |||||

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQNGIKANFKIRHINIEDGSGVOLAD 180
   |||||

QY 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||
DB 181 HYQONTPTIGDGPVLLPDNHYLSTQSALSQKDPNEKRHDHMLLEFVTAAGTTHGMDELYK 238
   |||||

RESULT 7
Q8WP95
ID Q8WP95 PRELIMINARY; PRT; 238 AA.
AC Q8WP95;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013824; AAK02062.1; -.
DR EMBL; AY013821; AAK02059.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
```

```
SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;
Query Match 42.8%; Score 1096; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.6e-70;
Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGHFKFSVSGEGGATYKGLTKFKICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGIVPVLIELDGDVGHGKFSVRGEGGADYKLEIKFKICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPGEYVQERTIFFDDGNGYKTRAEVKFEGDTLV 120
Db 61 VTTSYGIQCFARYPEHMKWNDFKFSAMPGEYIQERTIFFDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNNGKLVNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLLPDNHYLSQSALSKDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLLPINHYLSQTAISKDRNETRDHMFLEFFSACGHTHGMDELYK 238
RESULT 8
Q8WTC4
ID Q8WTC4 PRELIMINARY; PRT; 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435433; AAL3918.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;
Query Match 42.7%; Score 1093; DB 5; Length 238;
Best Local Similarity 83.6%; Pred. No. 2.6e-70;
Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGHFKFSVSGEGGATYKGLTKFKICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGVVPIVLVDGNGHFKFSVRGEGGADYKLEIKFKICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPGEYVQERTIFFDDGNGYKTRAEVKFEGDTLV 120
Db 61 VTTLGYILCFARYPEHMKWNDFKFSAMPGEYIQERTIFFDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNNGKLVNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLLPDNHYLSQSALSKDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLLPINHYLSQTAISKDRNETRDHMFLEFFSACGHTHGMDELYK 238
RESULT 9
Q8WTD0
ID Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL3913.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;
```

```
Q8WTD0
ID Q8WTD0 PRELIMINARY; PRT; 238 AA.
AC Q8WTD0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435427; AAL3912.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192173CB84D CRC64;
Query Match 42.6%; Score 1091; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 3.6e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLVDGNGHFKFSVSGEGGATYKGLTKFKICTTGGKLPVWPPTL 60
Db 1 MSKGEELFTGIVPVLIELDGDVGHGKFSVRGEGGADYKLEIKFKICTTGGKLPVWPPTL 60
QY 61 VTTLTYGVCFSRYPDHMKOHDFKFSAMPGEYVQERTIFFDDGNGYKTRAEVKFEGDTLV 120
Db 61 VTTLGYIQCFARYPEHMKWNDFKFSAMPGEYIQERTIFFDDGKYKTRGEVKFEGDTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGMDFKEDGNILGHKLEYNHNHNYIMPDKNNGKLVNFKIRHNIEGGVOLAD 180
QY 181 HYQONTPIGGPVLLPDNHYLSQSALSKDPNEKRDHMLVLEFVTAAGITHGMDLYK 238
Db 181 HYQTNVPLGDPVLLPINHYLSQTAISKDRNETRDHMFLEFFSACGHTHGMDELYK 238
RESULT 10
Q8WTC9
ID Q8WTC9 PRELIMINARY; PRT; 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN Aequorea macrodactyla.
OS Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435428; AAL3913.1; -
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;
```

```
Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVGHKFSVSGEGEGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVLDGVDVGHKFSVSGEGEGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 11
Q8WTC8 PRELIMINARY; PRT; 238 AA.
ID Q8WTC8 AC Q8WTC8 Q8WTC8
QY 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVGHKFSVSGEGEGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVLDGVDVGHKFSVSGEGEGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 12
Q8WTC7 PRELIMINARY; PRT; 238 AA.
ID Q8WTC7 AC Q8WTC7 Q8WTC7
QY 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435429; AAL33914.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVGHKFSVSGEGEGDATYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIVPLVLDGVDVGHKFSVSGEGEGDADYKGLKLFCTTGGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTLTGVCQFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNYKTRAEVKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDFKEDGNILGHKLEYNHNHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQONTPIGDPVLLPDNHYLSQTALSQKDPNEKRDHMLLEFVTAAGITHGMDELYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 13
Q8WTC5 PRELIMINARY; PRT; 238 AA.
ID Q8WTC5 AC Q8WTC5 Q8WTC5
QY 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPXM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF435432; AAL33917.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match 42.4%; Score 1086; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 8.1e-70;
```

Matches 198; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

```
QY 1 MSKGBELFTGVVPILVELDGDVNGHKFSYSGEGEDATYGLTKLKFTCTGKLPVPWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MSKGBELFTGVVPILVELDGDVNGHKFSYSGEGEDADYGLTKLKFTCTGKLPVPWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 VTTLYGYVOCFSSRYPDHMKQHDFFKFSAMPEGYVOERTIFFKDDGNKYKTRAEVAFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 VTTLYGYLCFARYPEHMKMDFKFSAMPEGYIOERTIFFQDDGKYKTRGEVAFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 NRIELKGIDFEDGNILGHKLEYWNSHNVYIMADKQNGIKANFKIRHNIEDGSVQLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 NRIELKGDMDFEDGNILGHKLEYWNSHNVYIMPDKANGLKYNFKIRHNIEGGVQLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 HYQNTPTIGDGPVLLPNHLYLSQSALSADPNKRDHMLVLEFVTAAGTHGMDELYK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 HYQNVPLGDGPVLPINLHLYLSQTSISKDRNETRDHMLVLEFVTAAGTHGMDELYK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 14

```
Q8WQY8 ID Q8WQY8 PRELIMINARY; PRT; 195 AA.
AC Q8WQY8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Apoeaquorin.
GN Aequorin.
OS Aequorea parva.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=148610;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AEQXXM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 3.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFh; 3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_3.
SQ SEQUENCE 195 AA; 22248 MW; 73EB72E7E0C140BE CRC64;
```

Query Match 35.7%; Score 915; DB 5; Length 195;
Best Local Similarity 85.3%; Pred. No. 9.7e-56;
Matches 162; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

```
QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 AVKLEPDPFNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 65
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 348 DAVEAFFGGAGMKYGVETDNPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 EAVEAFFGGAGMKYGVETENPEYIKGKKLAQTELDORFAKNQVTLIRLWGDALEFDIIDKD 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHIGFWYTMDDPA 467
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 QNGAITLDEWKYTLISAGIIQSAEDCEITFKVCDLDDSGRLDADEMTROHIGFWYTMDDPA 185
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 468 CEKLYGGAVP 477
   ||||| |||||
Db 186 CEKLYGGAVP 195
```

RESULT 15

```
Q8WQY7 ID Q8WQY7 PRELIMINARY; PRT; 195 AA.
AC Q8WQY7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
```

```
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Apoeaquorin.
GN Aequorin.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AEQXXM;
RA Luo W.X., Zhang J., Yang H.J., Li S.W., Xie X.Y., Qin Y.X., Pang S.Q.,
RA Li S.J., Xia N.S.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY013823; AAK02060.1; -.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 3.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_3.
SQ SEQUENCE 195 AA; 22266 MW; A459391B4B8003BF CRC64;
```

Query Match 35.1%; Score 899; DB 5; Length 195;
Best Local Similarity 83.7%; Pred. No. 1.4e-56;
Matches 159; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

```
QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 AVKLEPDPFNPRWIGRHKHMFNLDVNHNGKISLDEMVKASDIVINNLTGATPEQAKRHK 65
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 348 DAVEAFFGGAGMKYGVETDNPAYTEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 DAVEAFFGGAGMKYGVETENPEYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIIDKD 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETPRVCDIDESGOLDVDEMTROHIGFWYTMDDPA 467
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 QNGAITLDEWKYTLISAGIIQSAEDCEITFKVCDLDDSGRLDADEMTROHIGFWYTMDDPA 185
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 468 CEKLYGGAVP 477
   ||||| |||||
Db 186 CEKLYGGAVP 195
```

Search completed: March 13, 2003, 15:05:05
Job time : 50.0664 secs